Advances in Computational Solutions

Integrative Approaches and Applications

Editors

Dr. Ranjan Kumar

Abhishek Dhar

Dr. Ashes Banerjee

Arunima Mahapatra

Integrated Publications ™

New Delhi

Published By: *Integrated Publications* ™

Integrated Publications
H. No. - 3 Pocket - H34, Sector - 3,
Rohini, Delhi-110085, India
Email - info@integratedpublications.in

Editors: *Dr. Ranjan Kumar, Abhishek Dhar, Dr. Ashes Banerjee and Arunima Mahapatra*

Publication Year: *2024*

Edition: *1st*

Pages: *511*

ISBN: *978-93-5834-625-1*

Price: *₹2025/-*

Preface

In an era characterized by rapid advancements in technology and a deepening integration of computational methods into various scientific disciplines, the compilation of research presented in this book offers a window into the cutting-edge developments across a spectrum of fields. This collection embodies the essence of interdisciplinary innovation, merging theoretical foundations with practical applications to address complex problems and pioneer new solutions. Each chapter in this book delves into a unique aspect of modern research, reflecting the diversity and depth of current scientific inquiry: "Predictive Photo Responsivity of FASnBr3 Using Simulation" explores the photophysical properties of novel materials through advanced simulations, laying the groundwork for future applications in optoelectronics. "Optimizing Object Tracking Precision through Extended Kalman Filter Dynamics" provides a thorough examination of precision tracking technologies, enhancing our capability to monitor dynamic systems with unprecedented accuracy. "Modelling Error Factors in GNSS Observations for Precise Point Positioning and Accurate Position Estimation of Low Earth Orbit Satellites" addresses the critical need for precision in satellite positioning, which is vital for navigation and Earth observation. "Enhanced Performance in Multiple Object Tracking" combines Kalman Filters with the Hungarian Algorithm, optimizing tracking systems that are crucial in surveillance and autonomous navigation. "Computational Study of FeAsn (n=1-4) Clusters Invoking DFT-Based Descriptors" uses density functional theory to unravel the properties of iron-arsenic clusters, contributing to the field of materials science. "Automated Detection and Classification of Mass Calcifications in Breast Cancer Mammograms" leverages image processing techniques to advance the early detection of breast cancer, demonstrating the intersection of technology and healthcare. "The Emergence of Periodic and Chaotic Patterns in a Simple Predator-Prey System" investigates the dynamic behaviors in ecological models, enhancing our understanding of natural systems. "Dispersion Equation & Its Solution for Single Layer Fluid with Surface Tension" presents an analytical solution to a classic fluid dynamics problem, advancing theoretical fluid mechanics. "A Study on Internal Wave Velocity in a Two-Layer Fluid Medium" explores wave dynamics in stratified fluids, with implications for

oceanography and environmental sciences. "Formation of a Weakly Singular Integral Equation in Water Wave Scattering Problems" tackles mathematical challenges in modeling water wave interactions, providing numerical solutions using Bernstein polynomials. "Revisited Study on Advection Dominated Accretion Flow around Black Holes" revises existing models of accretion flows, contributing to our understanding of astrophysical phenomena. "Retailer's Inventory Decisions for Deteriorating Items" examines inventory management strategies, integrating the effects of advertisement and shortages, relevant for business logistics. "Impact of Carbon Emission on EPQ Model in Rhino Bricks Production" discusses environmental impacts on production models, aligning industrial processes with sustainability goals. "AR3D Face Recognition: A New Frontier in Human-Computer Interaction" explores advancements in biometric identification, pushing the boundaries of security and user interaction. "Augmented Reality and Virtual Reality in Education" highlights the transformative potential of immersive technologies in educational environments. "Chain of Wellness: Transforming Healthcare with Blockchain-Based Electronic Health Records" examines the role of blockchain in enhancing the security and efficiency of healthcare records management. "Deep Learning-Powered Brain Tumour Detection in MRI Imaging" demonstrates the application of deep learning in medical diagnostics, offering promising tools for healthcare professionals. "ChromaNet: Next-Gen Image Colorization with NoGAN" introduces innovative techniques in image processing, enhancing visual media through artificial intelligence. "Expense Forecasting on Tourism" applies machine learning to predict tourism expenses, aiding in economic planning and resource allocation. "IPL Player Auction Price Forecasting through Multiple Linear Regression" brings statistical analysis to sports management, predicting auction prices in cricket leagues. "Tip Trend: A Machine Learning Approach to Waiter Tip Prediction" uses data analytics to understand tipping behaviors, with practical applications in the hospitality industry. "Harnessing Adaboosting Algorithm for Predictive Money Management" explores financial forecasting models, contributing to better decision-making in finance. "Revolutionizing Employee Job Performance Assessment with Decision Tree Classification" offers innovative methods for evaluating job performance, enhancing human resource management. Each chapter represents a significant contribution to its respective field, embodying the

collective effort of researchers dedicated to pushing the boundaries of knowledge and application. As you embark on this journey through the pages of this book, we hope you find inspiration, insight, and a deeper appreciation for the remarkable advancements shaping our world today.

(Dr. Ranjan Kumar)

Associate Professor, Swami Vivekananda University,

Kolkata, West Bengal, India

Acknowledgement

I extend my heartfelt gratitude to Swami Vivekananda University, Kolkata, India, for their steadfast support and encouragement throughout the creation of "Advances in Computational Solutions: Integrative Approaches and Applications." The university's dedication to fostering education and research has been instrumental in shaping the content and direction of this publication. We deeply appreciate the collaborative spirit and resources provided by Swami Vivekananda University, Kolkata, which have enabled us to explore and share the latest innovations and technologies across various fields.

We hope that this book serves as a valuable resource for this esteemed institution and the broader academic community, reflecting our shared dedication to knowledge, progress, and the pursuit of excellence.

I extend my deepest appreciation to each of the external reviewers mentioned below for their unwavering commitment to excellence and their indispensable role in ensuring the scholarly merit of this work.

With sincere appreciation,

List of Reviewers

1. Prof. (Dr.) Subhas Chandra Mukhopadhyay, Mechanical and Electronics Engineering, Macquarie University, Australia
2. Dr. Arpan Deasi, Associate Professor, Department of ECE, RCCIIT, Kolkata
3. Prof. (Dr.) Himadri Biswas, Professor and Head, Dept. of CSE, Dr. Sudhir Chandra Sur Institute of Technology and Sports Complex, 540 Dum Dum Road, Surermath, Kolkata-74, WB
4. Dr. Debabrta Sarddar, Assistant Professor, Dept. of CSE, University of Kalyani, Kalyani, Nadia, PIN-741236
5. Dr. Payal Bose, Assistant Professor, Swami Vivekananda University, Barrackpore
6. Dr. Vassilis C. Gerogiannis, Professor, General Department of Larissa, University of Thessaly, Greece
7. Dr. Andreas Kanavos, Associate Professor, Department of Informatics, Ionian University

(Dr. Ranjan Kumar)

Associate Professor, Swami Vivekananda University,

Kolkata, West Bengal, India

Contents

Chapter - 1

Predictive Photo Responsivity of FASnBr3using Simulation

Authors

Saurabh Basak
Budge Budge Institute of Technology, Budge Budge, South 24 Parganas, Kolkata, West Bengal, India

Debraj Chakraborty
Swami Vivekananda University, Barrackpore, North 24 Parganas, Kolkata, West Bengal, India

Chapter - 1

Predictive Photo Responsivity of FASnBr3using Simulation

Saurabh Basak and Debraj Chakraborty

Abstract

Today organic tin-based perovskite solar cell is a new emerging technology. It has been proven by different experimental research that the best suited candidate of this group of perovskite is FASnI3. The experimental data of external quantum efficiency, integrated short circuit current density, photo responsivity is easily available in different research papers. This X site anion i.e iodine can be replaced by bromine (Br), chlorine (Cl) also but no such experimental data of the above said parameters of FASnX3 (X = Cl, Br, I) is available in literature. Since bromine (Br) is placed just above iodine (I) in the same halogen group we believe that it may have some similarity in photo responsivity and other parameters of FASnI3 based solar cells. So, in this research we propose a photo responsivity variation curveof FASnBr3 with respect to frequency of the incident light by simulation.

Keywords: Photoresponsivity, $FASnBr_3$, perovskite, solar cell.

Introduction

Perovskite based solar cells are new emerging technology for solar cell. Among them tin based organic perovskite solar cells are very suitable candidate for research due to easy synthesis process and reasonably high photo conversion efficiency which is increasing day by day. Kojima *et al* first invented this perovskite based solar cell in 2009 having photo conversion efficiency 3.8% [1-5]. Since then photo conversion efficiency has been increased to more than 25% [9-17] by today. But this perovskites are lead (Pb) based which causes several environmental hazards. So, tin based organic perovskite based solar cells came to picture to replace the former perovskite based solar cell. After a huge research on tin based organic perovskite solar cell the photo conversion efficiency becomes 14% for FASnI3 [2, 6-8]. The halogen iodine (I) can be replaced by other halogens.

In this research paper we place X = Br as halogen so our perovskite becomes FASnBr3. Since there are no such research paper showing the variation of external quantum efficiency (EQE) and photo responsivity (R) with respect to frequency of incident light for FASnBr3 we predict the above said variation by simulation.

Methodology

It is known to us that for any semiconductor light absorption takes place only when the light energy $E \geq E_g$, where E_g is the band gap of the semiconductor. Otherwise, it can be stated that if the frequency of the incident light $f \geq f_c$, where f_c is the cut off frequency of light absorption of that semiconductor then only absorption occurs otherwise this semiconductor will be acting as window layer for that semiconductor. So, if $f < f_c$ then EQE = η = 0 and photo responsivity = R = 0.

Cut off frequency f_c can be written as $f_c = \frac{E_g}{h}$, where h = Planck's constant. For $FASnI_3$ the E_g = 1.41 ev, $f_c = 333\ THz$ & maximum value of EQE = η = 76% at f = 857 THz [2]. The variation of EQE is from 60% - 76% within the frequency range of 375 - 857 THz. EQE decreases sharply after 857 THz due to surface absorption.

The algorithm of finding EQE of $FASnI_3$ is given below.

Step 1: Start.

Step 2: Read the frequency in THz (*f*)

Step 3: Check if $f < f_c$, η = 0, then else if *f < 375, η = 0.2,* then else if *f < 428.57,* η *= 0.7,* then else if *f < 858, η = 0.8,* else η *= 0.8exp(-100*1e-5*f).*

Step 4: Stop.

For $FASnBr_3$ band gap = 2.55 ev [18-21]. So, f_c = *615.75* THz.

From the EQE of $FASnI_3$ it is clear that after *f > 900* THz the wavelength become so small that it will be absorbed totally within the surface.

So, for higher value of *f*, the EQE and R will be sharply decreased.

As the electronegativity of Br is higher than I so for $FASnBr_3$, η must be lesser than $FASnI_3$.

This 100*1e-5 is the fitting parameter for EQE vs *f* curve of $FASnI_3$.

Result and discussion

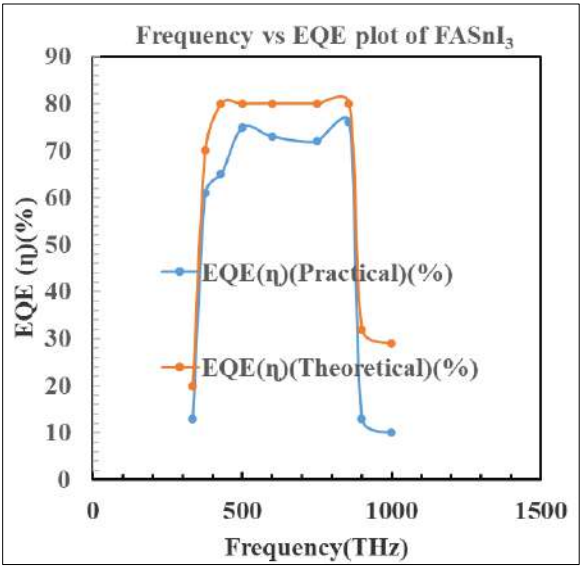


Figure 1: Frequency vs EQE plot of $FASnI_3$

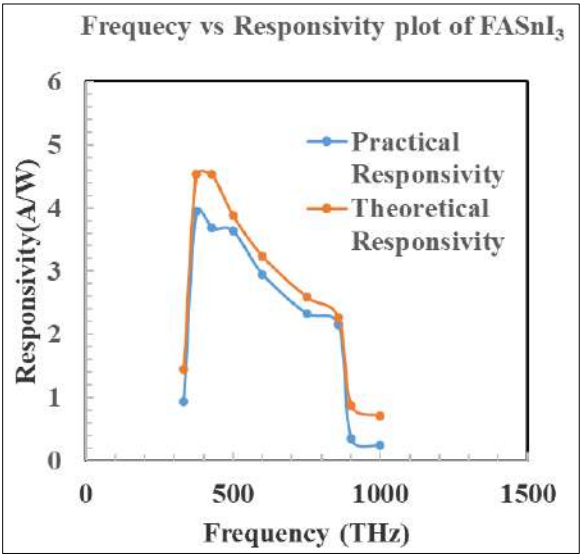


Figure 2: Frequency vs Responsivity plot of $FASnI_3$

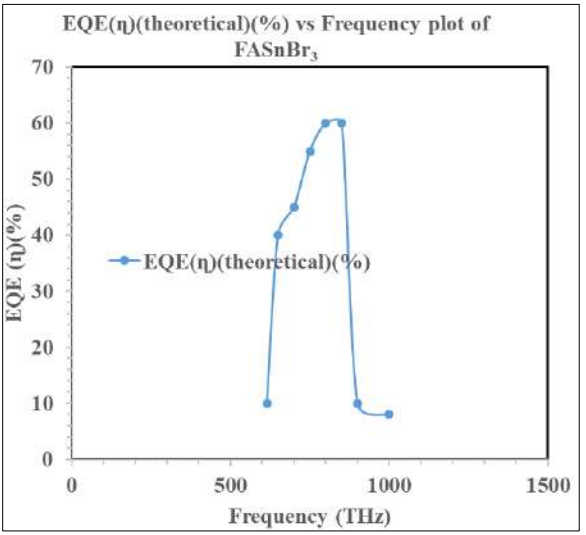


Figure 3: Frequency vs EQE (theoretical) plot of $FASnBr_3$

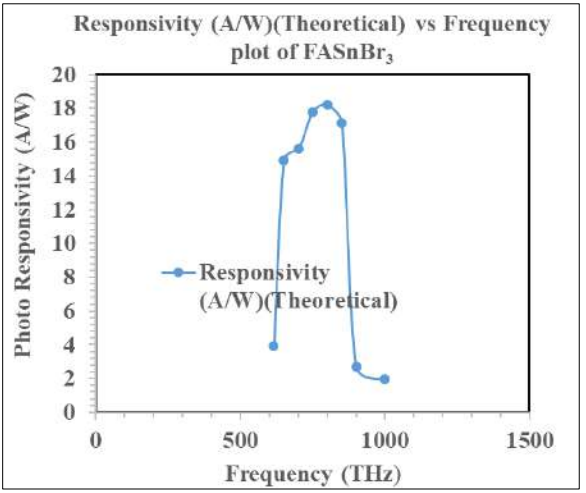


Figure 4: Frequency vs Responsivity (theoretical) plot of $FASnBr_3$

Fig.1 shows the variation of EQE w.r.t frequency of incident light of $FASnI_3$ for theoretical and practical both cases. Practical data of EQE w.r.t frequency variation of $FASnI_3$ is available in literature. From the formula $R = \frac{\eta q}{hf}$, we get the required photo responsivity value (R). In this case q = $1.6*10^{-19}$ Coulomb. Fig.2 shows the frequency vs Responsivity plot of $FASnI_3$ both theoretical & practical case.

In case of $FASnBr_3$, E_g = 2.55 ev, f_c = 615.75 THz. The frequency vs EQE (theoretical) plot of $FASnBr_3$ is given here in fig.3 by applying the above mentioned alogorithm with revised parametric values. The values of photo responsivity (R) is obtained by the above mentioned formula and plotted w.r.t frequency of the incident light.

Conclusion

The predicted external quantum efficiency (EQE) and photo responsivity (R) curve of $FASnBr_3$ has been plotted here. It will help us for further study related to $FASnBr_3$ based perovskite solar cell.

References

1. Fang Hong - Hua, Adjokatse Sampson, Shao Shuyan, Even Jacky, Loi Maria Antonietta, (2018) "Long - lived hot - carrier light emission and large blue shift in formamidinium tin triiodide perovskites". *Nature communications*, 9 (243): 243.
2. Zhu Zihao, Jiang Xianyuan, Yu Danni, Yu Na, Ning Zhijun, Mi Qixi,(2022) " Smooth and Compact $FASnI_3$ films for Lead - Free Perovskite Solar Cells with over 14% Efficiency" *ACS Energy Lett.*,7, 2079-2083
3. Baig F., Khattak Y.H., Mari B., Beg S., Ahmed A., Khan K., (2018) "Efficiency Enhancement of $CH_3NH_3SnI_3$ Solar Cells by Device Modelling," *J. Electron. Mater*. 47 (9), 5275-5282.
4. Farooq W., Tu S., Iqbal K., Khan H. A., Rehman S. U., Khan A. D., Rehman O. U.,(2020) "An Efficient Non -Toxic and Non - Corrosive Perovskite Solar Cell" *IEEE Access*, 8, 210617 - 210625.
5. Moiz S. A., Alahmadi A.N.M., Aljohani A.J., (2021) "Design of a Novel Lead - Free Perovskite Solar Cell For 17.83% Efficiency" *IEEE Access*, 8, 54254 - 54263.
6. Nasti G., Aldamasy M.H., Flatken M.A., Musto P., Matczak P., Dallmann A., Hoell A., Musiienko A., Hempel H., Girolamo E.A.D.D., Pascual J., Li G., Li M., Mercaldo L.V., Veneri P.D., and Abate A., (2022) " Pyridine Controlled Tin Perovskite Crystallization" *ACS Energy Lett*., 7, 3197−3203.
7. Meng X., Wu T., Liu X., He X., Noda T., Wang Y., Segawa H., Han L., (2020) "Highly Reproducible and Efficient $FASnI_3$ Solar Cells Fabricated with Volatilizable Reducing Solvent" *J. Phys. Chem. Lett.*, 11, 2965 - 2971.

8. Nakamura T., Handa T., Murdey R., Kanemitsu Y., Wakamiya A., (2020) "Materials Chemistry Approach for Efficient Lead - Free Tin Halide Perovskite Solar Cells." *ACS Appl. Electron. Mater.*, 2, 3794 - 3804.

9. Wang M., Wang W., Ma B., Shen W., Liu L., Cao K., Chen S., Huang W., (2021) "Lead-Free Perovskite Materials for Solar Cells" *Nano-Micro Lett*, 13:62,

10. Li M., Li F., Gong J., Zhang T., Gao F., Zhang W. H., Liu M., (2021) "Advances in Tin(II)-Based Perovskite Solar Cells: From Material Physics to Device Performance" *small struct.*, 2100102, 1 - 30.

11. Mutalib M.A., Ludin N.A., Ruzalman N.A.A.N., Barrioz V., *et al.*(2018) "Progress towards highly stable and lead-free perovskite solar cells" *Materials for Renewable and Sustainable Energy*, 7:7.

12. Boix P.P., Nonomura K., Mathews N., Mhaisalkar S.G., (2014) "Current progress and future perspectives for organic/inorganic perovskite solar cells". *Mater. Today*, 17, 16 -23

13. Green M.A., Ho-Baillie A., Snaith H.J., (2014) "The emergence of perovskite solar cells". *Nat Phot,* 2014, 8, 506-514.

14. Baranwal A.K., Hayase, S., (2022) "Recent Advancements in Tin Halide Perovskite-Based Solar Cells and Thermoelectric Devices" *Nanomaterials*, 12, 4055.

15. Green M.A., Baillie A.H., Snaith A.J., (2014) "The emergence of perovskite solar cell" *Nature Photonics*, Vol 8.

16. Li C., *et al.* (2008) "Formability of ABX_3 (X = F, Cl, Br, I) halide perovskites". *Acta Crystallogr.* B 64, 702-707.

17. Wang L., Ou T., Wang K., Xiao G., Gao C., Zou B., (2017) " Pressure-induced structural evolution optical and electronic transitions of nontoxic organometal halide perovskite-based methylammonium tin chloride". Appl. Phys. Lett. 111, 233901.

18. Lu C. H., McGee G.B., Liu Y., Kang Z., Lin Z., (2020) "Doping and Ion Substitution in Colloidal Metal Halide Perovskite Nanocrystals". *Chemical Society Reviews* 1-130.

19. Das T., Liberto G.D., Pacchioni G., (2022) "Density Functional Theory Estimate of Halide Perovskite Band Gap: When Spin Orbit Coupling Helps". *J. Phys. Chem. C*, 126, 2184-2198.

20. Zhou F.; Qin F.; Yi Z.; Yao W.; Liu Z.; Wu X.; Wu P. (2021) "Ultra-wideband and wide-angle perfect solar energy absorber based on Ti nanorings surface plasman resonance". *Phys. Chem. Chem. Phys*., 23, 17041.

21. Saliba M., Matsui T., Domanski K., Seo J.-Y., Ummadisingu A., Zakeeruddin S.M., Correa-Baena, J.-P., Tress,W.R., Abate A., Hagfeldt A., *et al*., (2016) " Incorporation of rubidium cations into perovskite solar cells improves photovoltaic performance". *Science*, 354, 206-209.

Chapter - 2

Optimizing Object Tracking Precision through Extended Kalman Filter Dynamics: A Comprehensive Investigation and Performance Evaluation

Author

Sk Babul Akhtar
Swami Vivekananda University, Barrackpore, North 24 Parganas, Kolkata, West Bengal, India

Chapter - 2

Optimizing Object Tracking Precision through Extended Kalman Filter Dynamics: A Comprehensive Investigation and Performance Evaluation

Sk Babul Akhtar

Abstract

The field of object tracking has witnessed significant research over the past decade, presenting numerous challenges related to understanding the system model governing object movement and the type of noise it encounters. These challenges involve distinguishing between linear or nonlinear system models, such as the coordinated turn model, and addressing Gaussian or non-Gaussian noise, depending on the chosen filtering approach. The Extended Kalman Filter (EKF) has gained widespread use in tracking dynamic objects like missiles, aircraft, and robots. Considering two distinct models, this study focuses on a Bearing-Only Tracking (BOT) problem with a single sensor or observer. In the first model, the target is assumed to exhibit constant velocity and a consistent course, while the second model assumes a coordinated turn with constant velocity but variable course. The Extended Kalman Filter is employed to track the target in both scenarios effectively. For certain applications, incorporating elements of nonlinearity and non-Gaussianity becomes imperative to accurately model the underlying dynamics of a system. This research employs and compares the performance of the nonlinear Extended Kalman Filter (EKF) and the particle filter (PF) algorithms in tracking maneuvering objects with bearing-only measurements. A systematic approach to resampling particles is proposed to mitigate the degeneracy effect in the particle filter (PF) algorithm, with a specific focus on the sequential importance resampling (SIR) technique. The efficacy of this technique is discussed and compared with the standard EKF through an illustrative example.

Introduction

Tracking involves determining the location of an object over time.

Typically, the tracking process begins by detecting the object, followed by tracking its movement, and ultimately observing its behavior as it unfolds over time. Object tracking is essential in numerous real-world applications, including surveillance, video compression, and video analysis. Throughout the object's trajectory, various challenges arise, prompting the application of different tracking techniques to address and overcome them [1, 2].

A filter, in this context, refers to a mechanism that accepts specific data as input, processes and transforms it, and then produces the transformed output. In the context of tracking an object's trajectory, the utilization of filters arises from the presence of uncertainties during the tracking process. These uncertainties include noise, occlusions, and changes in appearance along the object's path. Filtering serves as a technique to mitigate the impact of such uncertainties, particularly by eliminating the existing noise [3]. The concept of filtering essentially involves the method of isolating and refining estimations, ensuring an optimal assessment of the state. When applied to object tracking, filters play a crucial role in removing undesirable noises, hence the term “tracking with filters” [4]. Several models are utilized in object tracking, with some widely employed ones discussed below:

Theory

1. Constant velocity model: In this model, the object moves with a constant velocity, implying negligible acceleration throughout. Acceleration terms are not considered during each time step. The present x and y positions are determined by adding the velocity in the respective directions to the previous time step's positions [5, 6].
2. Constant acceleration model: In the constant acceleration model, the object maintains a constant acceleration, and this acceleration term is incorporated during the update of x and y coordinates at each time instant [7-9]. The present positions are calculated based on the previous positions, velocities, and acceleration terms.
3. Coordinated turn model: The coordinated turn model introduces a new term known as the coordinated turn rate, denoted as "w" and measured in radians per second. While updating the x and y coordinates, this model considers the velocity term but excludes the acceleration term, incorporating the newly added coordinated turn rate [10].

The constant velocity model is considered linear since it relies on previous positions and velocities, resembling a linear model. The constant

acceleration model, although not strictly linear due to the inclusion of acceleration terms, is not highly nonlinear. On the other hand, the coordinated turn model is nonlinear, as it includes both the constant velocity term and the coordinated turn rate during coordinate updates.

Extended kalman filter

The extended Kalman filter broadens the applicability of the traditional Kalman filter [11, 12] to address nonlinear state estimation challenges. This is achieved by constructing a Gaussian approximation for the joint distribution of the state variable "x" and measurements "y" through a transformation based on the Taylor series. The filtering model employed in the extended Kalman filter is:

$$X_{k+1} = f(X_k, k) + w_k \quad (1)$$

$$z_k = h(X_k, k) + v_k \quad (2)$$

Here X_k denotes the state vector at kth time step w_k is the process noise vector v_k is measurement noise vector $f(X_k, k)$ is nonlinear system state transition function which transits state from one time instant to another time instant $h(X_k, k)$ is measurement function z_k is measurement state vector at kth time instant. Much like the Kalman filter, the extended Kalman filter follows a two-step process to estimate unknown states or variables [11-15]. These steps are:

1. Prediction: In the prediction step, a forecast of the new state vector and its corresponding covariance is computed over a specific time interval. This is achieved by utilizing dynamic system and measurement models. Following the prediction, residual measurement error and residual error covariance matrices are calculated. These values are then employed in the subsequent state and covariance updating process, aiding in the determination of the Kalman gain.
2. Updating: After predicting the state vector and covariance matrices, the next step involves updating these values. The process includes the calculation of the Kalman gain matrix. Once the Kalman gain is obtained, the final estimation of the state vector and covariance matrices is accomplished, incorporating this Kalman gain term into the calculations [8, 12].

Prediction:

$$X_k = f(X_{k-1}, u_{k-1}, 0) \tag{3}$$

$$P_k = A_k P_{k-1}, A_k^T + W_k Q_{k-1} W_k^T \tag{4}$$

Kalman Gain:

$$K_k = P_k H_k^T (H_k P_k H_k^T + v_k R_k v_k^T)^{-1} \tag{5}$$

Updation:

$$P1_k = (I - K_k H_k) P_k \tag{6}$$

$$X1_k = X_k + K_k (Z_k - h(X_k, 0)) \tag{7}$$

Here H_k is the measurement Jacobian matrix, A_k is the Jacobian for state transition, Q_{k-1} is the process noise covariance matrix, R_k is the measurement noise covariance matrix.

Methodology

When monitoring the movement of a specific object equipped with surrounding sensors in a defined area, the sensors provide bearing angles to the object [16]. These bearing angles are determined by referencing the object's positions with the fixed coordinates of the sensors. Typically, sensors maintain fixed coordinates, and as time progresses from one-time step to the next, they continuously provide bearing angles based on the object's movement within the specified area.

The nonlinearity in the measurement model specifically pertains to the bearing when tracking. For instance, if a sensor is stationed at the origin and remains fixed throughout the experiment, the exploration of bearing angles is discussed in [1, 3, 20].

$$h1_k = tan^{-1}\left(\frac{y_k - y_s}{x_k - x_s}\right) \tag{8}$$

The Jacobian of this measurement model is evaluated by calculating the partial derivative of $h1_k$ with respect to state variables [16, 17].

$$H_k = \left[\frac{-(y_k - y_s)}{(y_k - y_s)^2 + (x_k - x_s)^2} \quad \frac{(x_k - x_s)}{(y_k - y_s)^2 + (x_k - x_s)^2} \quad 0 \; 0\right] \tag{9}$$

Considering multiple measurements,

$$H_k = \left[\frac{-(y_k-y_s^1)}{(y_k-y_s^1)^2+(x_k-x_s^1)^2} \;\; \frac{(x_k-x_s^1)}{(y_k-y_s^1)^2+(x_k-x_s^1)^2} \;\; 0\;0 \;\; \frac{-(y_k-y_s^2)}{(y_k-y_s^2)^2+(x_k-x_s^2)^2} \;\; \frac{(x_k-x_s^1)}{(y_k-y_s^2)^2+(x_k-x_s^2)^2} \;\; 0\;0 \;\; \frac{-(y_k-y_s^1)}{(y_k-y_s^3)^2+(x_k-x_s^3)^2} \;\; \frac{(x_k-x_s^1)}{(y_k-y_s^3)^2+(x_k-x_s^3)^2} \;\; 0\;0 \;\; \frac{-(y_k-y_s^1)}{(y_k-y_s^4)^2+(x_k-x_s^4)^2} \;\; \frac{(x_k-x_s^1)}{(y_k-y_s^4)^2+(x_k-x_s^4)^2} \;\; 0\;0\right] \tag{10}$$

If we take into consideration the Hessian parameters [17- 19] as well, the H matrix becomes:

$$H_k = \left[\frac{-2(x_k-x_s^n)}{((y_k-y_s^n)^2+(x_k-x_s^n)^2)^2} \;\; \frac{(y_k-y_s^n)^2-(x_k-x_s^n)^2}{((y_k-y_s^n)^2+(x_k-x_s^n)^2)^2} \;\; 0\;0 \;\; \frac{(y_k-y_s^n)^2-(x_k-x_s^n)^2}{((y_k-y_s^n)^2+(x_k-x_s^n)^2)^2} \;\; \frac{-2(y_k-y_s^n)}{((y_k-y_s^n)^2+(x_k-x_s^n)^2)^2} \;\; 0\;0\;0\;0\;0\;0\;0\;0\;0\;0\right] \tag{11}$$

Constant velocity model

In consistent velocity model object is moving with constant velocity for this model state transition matrix can be written as:

$$F = [1\;0\;t\;0\;0\;1\;0\;t\;0\;0\;1\;0\;0\;0\;0\;1\;] \tag{12}$$

Let's focus on the state vector specific to bearing-only tracking in this model, denoted as $X = [x, y, \dot{x}, \dot{y}]^T$. Here, $'x'$ and $'y'$ signify the positions along the x and y axes, respectively, while $\dot{x}$ and $\dot{y}$ represent the velocity components corresponding to the x and y directions, respectively.

In the context of bearing-only tracking, envision a scenario where an observer is situated at a specific point. Unlike a sensor, this observer has fixed x and y coordinates, maintaining a constant position throughout the experiment. The location of this observer is denoted as (x_s, y_s).

Constant acceleration model

In the context of the constant acceleration model, the object maintains a consistent acceleration throughout the entire time interval of the experiment. This results in an expansion of the state variables, involving the inclusion of acceleration terms in both the x and y directions to the existing position and velocity terms.

The modified state vector is represented as:

$$X = [x, y, \dot{x}, \dot{y}, \ddot{x}, \ddot{y}]^T \quad (13)$$

In the constant acceleration model, the state transition matrix is characterized by:

$$F = \left[1\ 0\ t\ 0\ \frac{t^2}{2}\ 0\ 0\ 1\ 0\ t\ 0\ \frac{t^2}{2}\ 0\ 0\ 1\ 0\ t\ 0\ 0\ 0\ 0\ 1\ 0\ t\ 0\ 0\ 0\ 0\ 1\ 0\ 0\ 0\ 0\ 0\ 0\ 1\right] \quad (14)$$

where t is the sampling interval.

Constant turn model

In the coordinated turn model, an object maintains a constant velocity while simultaneously altering its course or direction over time. The term "course" here is synonymous with direction, signifying that the object undergoes continuous changes in its orientation, essentially turning as time progresses. In this model, the state vector undergoes augmentation with the inclusion of a new parameter: the coordinated turn rate, denoted as $WWkk$ and measured in radians per second. The coordinated turn rate is also referred to as the angular turn rate.

In the coordinated turn model, we assume the object follows the constant velocity model within a two-dimensional system. Notably, acceleration terms are omitted from this model, as the object is considered to maintain a constant velocity, implying zero rate of change in velocity. Consequently, the state vector for this model is structured as follows [3]:

$$X = [x, y, \dot{x}, \dot{y}, w]^T \quad (15)$$

For the above coordinated turn model state transition matrix is explored in [5].

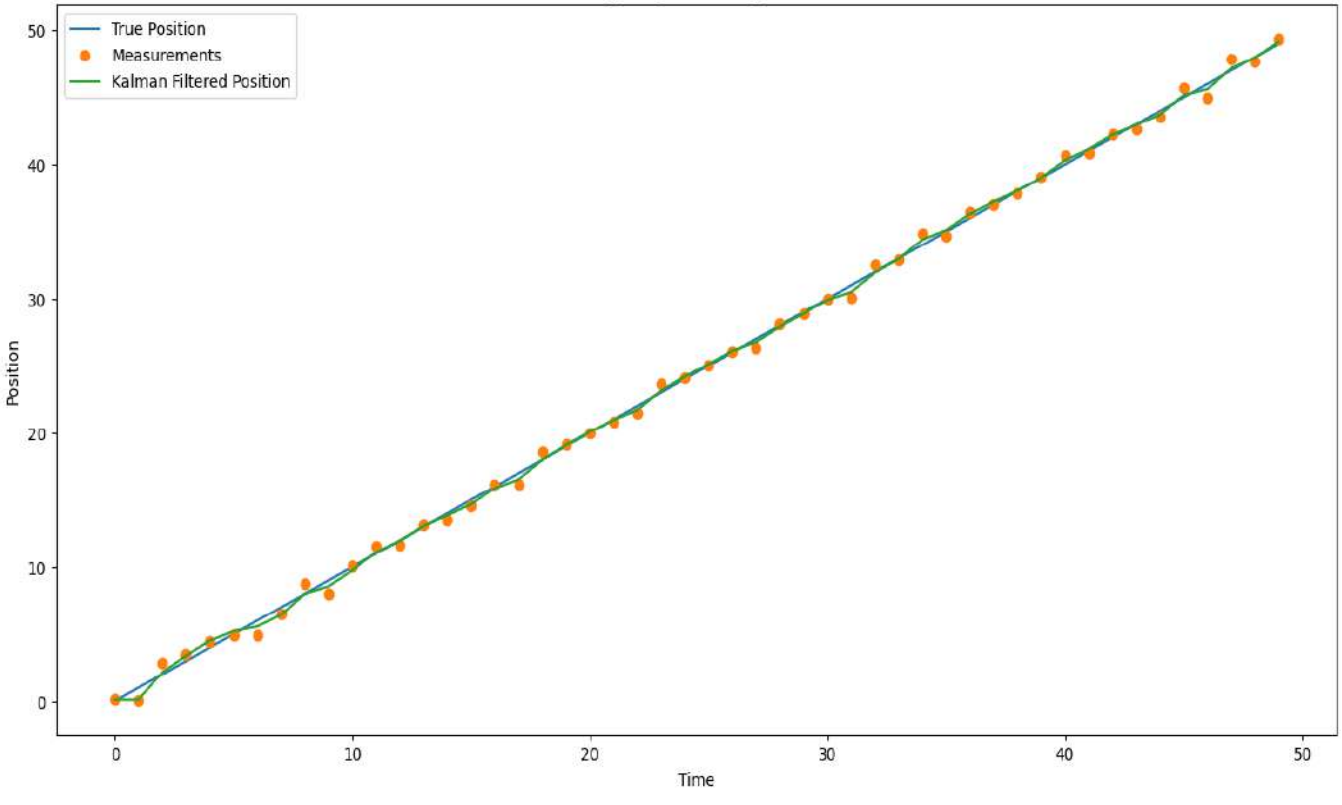
$$F = \left[1\ 0\ \frac{sin(w\Delta k)}{w}\ \frac{coscos\,(w\Delta k)-1}{w}\ \frac{\partial x_{t+1}}{\partial w}\ 0\ 1\ \frac{1-cos(w\Delta k)}{w}\ \frac{sinsin\,(w\Delta k)-1}{w}\ \frac{\partial y_{t+1}}{\partial w}\ 0\ 0\ sin(w\Delta k)\ -\ sin(w\Delta k)\ \frac{\partial \dot{x}_{t+1}}{\partial w}\ 0\ 0\ cos(w\Delta k)\ cos(w\Delta k)\ \frac{\partial \dot{y}_{t+1}}{\partial w}\ 0\ 0\ 0\ 0\ 1\right] \quad (16)$$

where Δk is the sampling time interval.

Result and analysis

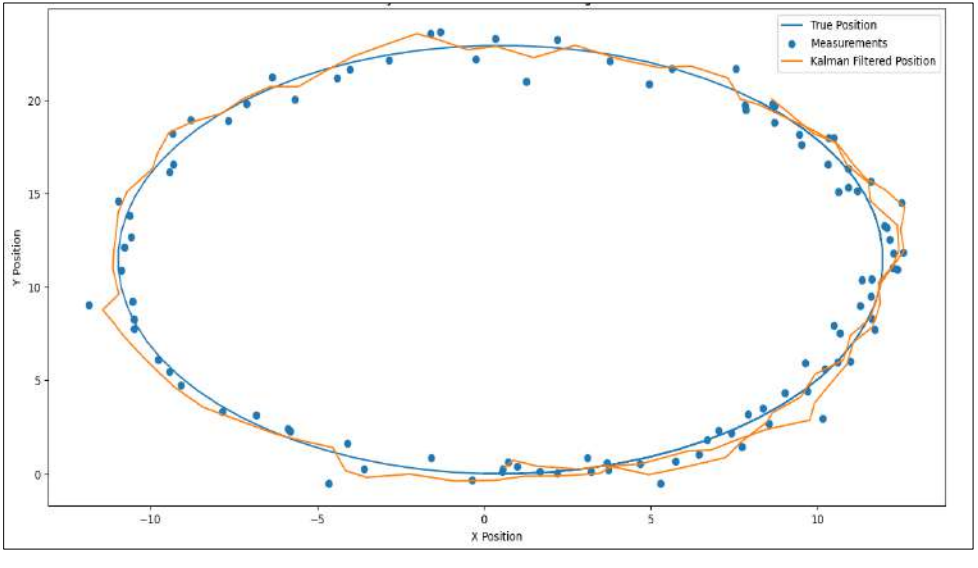
Constant velocity and course target

Referring to Figure 1, when the process model is linear and the noise distribution adheres to a Gaussian distribution, assuming an accurate initial condition, the Kalman filter adeptly tracks the object's path. In this scenario, the object commences its journey at the initial coordinates maintaining a constant speed.


True Position
Measurements
Kalman Filtered Position
Position
Time
0
10
20
30
40
50

Constant velocity and varying course target

Examining Figure 2, it illustrates the tracking of a maneuvering target's trajectory using the Extended Kalman Filter with a coordinated turn model, incorporating a particular turn rate [5]. The target initiates its movement from the initial coordinates while maintaining a constant velocity. The simulation is repeated, considering the nonlinear characteristics of the coordinate turn model. The assumed noise follows a Gaussian distribution, and the Extended Kalman Filter effectively traces the object's path, providing optimal estimations.



Future plan and conclusion

The analysis of target tracking using the Extended Kalman Filter has yielded significant insights from various scenarios. Notably, the Kalman filter demonstrates superior performance when the system model adheres to linearity and the assumed noise distribution follows a Gaussian pattern. On the other hand, the Extended Kalman Filter proves effective even in cases where the system model is nonlinear, provided the noise distribution remains Gaussian. However, when dealing with heightened nonlinearity or non-Gaussian distribution scenarios, the performance of the Extended Kalman Filter diminishes. In such instances, particle filters emerge as a more optimal choice for precise tracking.

Looking ahead, future work could delve into refining the Extended Kalman Filter's adaptability to handle increased nonlinearity and non-Gaussian noise distributions. Additionally, exploring enhancements or alternative methodologies that address these challenges in diverse tracking scenarios would contribute to the advancement of target tracking systems.

References

1. Pal, Sankar K., *et al*. "Deep learning in multi-object detection and tracking: state of the art." Applied Intelligence 51 (2021): 6400-6429.
2. Zhang, Yifu, *et al*. "Bytetrack: Multi-object tracking by associating every detection box." European Conference on Computer Vision. Cham: Springer Nature Switzerland, 2022.
3. Wang, Zhongdao, *et al*. "Towards real-time multi-object tracking." European Conference on Computer Vision. Cham: Springer International Publishing, 2020.
4. Bescos, Berta, *et al*. "DynaSLAM II: Tightly-coupled multi-object tracking and SLAM." IEEE robotics and automation letters 6.3 (2021): 5191-5198.
5. Tokmakov, Pavel, *et al*. "Learning to track with object permanence." Proceedings of the IEEE/CVF International Conference on Computer Vision. 2021.
6. Wu, Hai, *et al*. "3D multi-object tracking in point clouds based on prediction confidence-guided data association." IEEE Transactions on Intelligent Transportation Systems 23.6 (2021): 5668-5677.
7. Jiang, Rui, *et al*. "Object tracking on event cameras with offline-online

learning." CAAI Transactions on Intelligence Technology 5.3 (2020): 165-171.

8. Mauri, Antoine, *et al*. "Deep learning for real-time 3D multi-object detection, localisation, and tracking: Application to smart mobility." Sensors 20.2 (2020): 532.
9. Fang, Zheng, *et al*. "3d-siamrpn: An end-to-end learning method for real-time 3d single object tracking using raw point cloud." IEEE Sensors Journal 21.4 (2020): 4995-5011.
10. Zheng, Chaoda, *et al*. "Box-aware feature enhancement for single object tracking on point clouds." Proceedings of the IEEE/CVF International Conference on Computer Vision. 2021.
11. Zolfaghari, Mohammad, Hossein Ghanei-Yakhdan, and Mehran Yazdi. "Real-time object tracking based on an adaptive transition model and extended Kalman filter to handle full occlusion." The Visual Computer 36 (2020): 701-715.
12. Thormann, Kolja, Shishan Yang, and Marcus Baum. "A comparison of kalman filter-based approaches for elliptic extended object tracking." 2020 IEEE 23rd International Conference on Information Fusion (FUSION). IEEE, 2020.
13. Farahi, Fahime, and Hadi Sadoghi Yazdi. "Probabilistic Kalman filter for moving object tracking." Signal Processing: Image Communication 82 (2020): 115751.
14. Rashid, Ekbal, *et al*. "Improvement in extended object tracking with the vision-based algorithm." Modern Approaches in Machine Learning and Cognitive Science: A Walkthrough: Latest Trends in AI (2020): 237-245.
15. Kim, Taeklim, and Tae-Hyoung Park. "Extended Kalman filter (EKF) design for vehicle position tracking using reliability function of radar and lidar." Sensors 20.15 (2020): 4126.
16. Xu, Yanzhe, *et al*. "Improved small blob detection in 3D images using jointly constrained deep learning and Hessian analysis." Scientific reports 10.1 (2020): 326.
17. Wang, Weixing, *et al*. "Lane line extraction in raining weather images by ridge edge detection with improved MSR and Hessian matrix." Information Technology and Control 50.4 (2021): 722-735.

18. Rana, Md Masud, *et al*. "Position and velocity estimations of 2D-moving object using Kalman filter: Literature review." 2020 22nd International Conference on Advanced Communication Technology (ICACT). IEEE, 2020.
19. Farag, Wael. "Kalman-filter-based sensor fusion applied to road-objects detection and tracking for autonomous vehicles." Proceedings of the Institution of Mechanical Engineers, Part I: Journal of Systems and Control Engineering 235.7 (2021): 1125-1138.
20. Pallavi, S., *et al*. "Study and analysis of modified mean shift method and Kalman filter for moving object detection and tracking." Proceedings of the Third International Conference on Computational Intelligence and Informatics: ICCII 2018. Springer Singapore, 2020.

Chapter - 3

Modelling Error Factors in GNSS Observations for Precise Point Positioning and Accurate Position Estimation of Low Earth Orbit Satellites

Author

Sk Babul Akhtar
Swami Vivekananda University, Barrackpore, North 24 Parganas, Kolkata, West Bengal, India

Chapter - 3

Modelling Error Factors in GNSS Observations for Precise Point Positioning and Accurate Position Estimation of Low Earth Orbit Satellites

Sk Babul Akhtar

Abstract

This paper focuses on the comprehensive modelling of error factors influencing Global Navigation Satellite System (GNSS) observations in the context of Precise Point Positioning (PPP) for the accurate estimation of positions in Low Earth Orbit (LEO) satellites. The study delves into various sources of error, including satellite clock error, ephemerides prediction error, atmospheric delays (Ionospheric and Tropospheric Delays), receiver noise, Sagnac effect, and relativistic clock corrections. Each error source is meticulously examined to understand its impact on GNSS observations during the precise positioning. The proposed model aims to enhance position estimation accuracy by accounting for and mitigating these error factors. In the pursuit of precise point positioning, the paper outlines the methodology employed to address and correct these errors. Utilizing advanced techniques, including rigorous modelling and correction algorithms, the study achieves accurate positioning results for Low Earth Orbit Satellites. The findings demonstrate the effectiveness of the proposed error modelling approach, highlighting its significance in improving the precision and reliability of GNSS-based position estimation for satellite applications in LEO.

Introduction

In contemporary times, the GNSS-based orbit determination technique is extensively employed for determining the orbits of Low Earth Orbit (LEO) satellites, owing to its cost-effectiveness and easy applicability. GNSS enables autonomous global positioning, with ground or space-based receivers capturing signals from orbiting navigation satellites to ascertain their three-dimensional location and precise time. These receivers analyse both pseudo-range (code) and carrier phase data transmitted by GNSS

satellites. Since 1994, the International GNSS Service (IGS) has been fulfilling its mission to provide openly accessible high-precision GNSS data and products [1-3]. The accuracy of GNSS signals is influenced by various factors, including ephemeris errors, satellite and receiver clock discrepancies, atmospheric effects, multipath distortion, and more.

Global Navigation Satellite Systems (GNSS) encompass constellations of satellites orbiting Earth, transmitting spatial coordinates and delivering precise, continuous, global, three-dimensional position and velocity data to users equipped with suitable receiving devices. GNSS finds applications across various transportation modes, including space stations, aviation, rail, road, and mass transit. The crucial role of positioning, navigation, and timing extends to telecommunications, land surveying, emergency response, mining, scientific research, and other fields. GNSS satellites broadcast both the transmission time and their corresponding positions, utilizing two carrier waves in the L−Band at frequencies $L1 = 1575.42\,MHz$ and $L2 = 1227.60\,MHz$. These transmissions, known as the Navigation Message or Nav Message, contain crucial information, including an accurate timestamp derived from the satellites' highly precise atomic clocks [4]. Additionally, the Nav Message includes satellite ephemeris, providing the satellite's position. Upon receiving the signal, a GNSS receiver utilizes its internal clock to calculate the time of reception. The time difference between transmission and reception is then employed to determine the range, where the product of this time difference (Δt seconds) and the speed of light (299,792,458 m/s) yields the satellite's range. It is important to note that this range does not precisely represent the geometric distance from the satellite to the receiver due to inherent errors.

Factors affecting GNSS observations

GNSS provides two types of observations: code and carrier phase measurements. While carrier phase measurements offer higher accuracy compared to code phase measurements, they are susceptible to integer ambiguity [5, 6]. The presence of integer ambiguity can introduce uncertainty in the exact number of carrier wavelengths between the transmitter and receiver. Various factors contribute to the reduction in accuracy for both code and carrier phase observations, highlighting the complexity involved in achieving precise positioning in GNSS applications.

1. Satellite clock error: Satellite clock error arises from the bias and drift in GNSS satellite clocks, impacting the accuracy of the GNSS

signal for range correction. To achieve precise positioning, the International GPS Services (IGS) delivers precise estimations of satellite clock bias and drift, although real-time support is currently unavailable [7]. The master control station in Colorado Springs explicitly determines and monitors satellite clock bias, drift, and drift rate, broadcasting these parameters through the navigation message.

2. Ephemerides prediction error: It arises from deviations in the precise orbits of navigation satellites caused by gravitational forces. Regular monitoring and control of satellite positions are implemented to mitigate orbit data errors. Corrected data, including ephemeris information, is disseminated through the navigation message. The position of a transmitting GNSS satellite is derived by propagating the GNSS broadcast ephemerides. However, since ephemerides are not frequently updated, prolonged propagation can lead to position errors. To ensure accuracy in precise positioning, the use of the IGS ephemeris product is essential for error correction. Real-time reduction of this error is possible by employing an empirical propagation model [7, 8].
3. Atmospheric delays: Atmospheric delay introduces a delay in signal reception due to the influence of the atmosphere on electromagnetic wave propagation. This delay can be categorized into Tropospheric and Ionospheric delays. Tropospheric models, such as the Saastamoinen model, estimate delay caused by pressure, temperature, and humidity in the troposphere. The impact of the ionosphere is more pronounced for satellites closer to the horizon. Ionospheric delay can also be computed using a global ionospheric model [7, 10].
4. Ionospheric error correction: Using both L1 and L2 pseudo-range measurements ρ_{L1} and ρ_{L2}, the ionospheric error on L1 is estimated as [9, 11]:

$$\Delta S_{ion-corr,L1} = \left(\frac{f_{L2}^2}{f_{L2}^2 - f_{L1}^2}\right)(\rho_{L1} - \rho_{L2}) \quad (1)$$

Tropospheric delay correction: The tropospheric delay is estimated as [7, 11]:

$$S_{tropo} = m_d . d_{dry} + m_w . d_{wet} \quad (2)$$

where m_d = dry-component mapping function, m_w = wet-component mapping function, d_{dry}= dry components of the delay, d_{wet} = wet components of the delay As the LEO satellite is well above the troposphere, this error correction is not required in the present model.

5. Receiver noise: This is the result of thermal noise in the receiver and is considered random noise in the measurement model [12].

6. Sagnac Effect: In addition to the satellite's motion during the signal's flight time, the Earth's eastward rotation introduces an error of approximately 30 meters in the eastward direction. The calculation of the GPS satellite position at the time of signal reception (t_r) involves utilizing the broadcast ephemeris data to determine the satellite's position at the time of signal transmission (t_t) [9, 13].

$$r_{gps}(t_r) = R_z(\theta) r_{gps}(t_t) \quad (3)$$

During this time $\Delta t = t_r - t_t$ the Earth rotates at an angle $\theta = \omega_{\oplus} \Delta t$ about z-axis with an angular rate $\omega_{\oplus} = 7.2921151467 \times 10^{-5}\ rad/sec$

7. Relativistic clock correction and shapiro effect: The application of the post-Newtonian theory of relativity is essential for making relativistic corrections to Global Positioning System (GPS) observations [14]. Discrepancies between the internal clock of a Low Earth Orbit (LEO) satellite and a clock situated on the Earth's surface arise from both general relativity stemming from the gravitational potential disparity between the signal source and receiver—and special relativity resulting from the relative speed between a moving emitter and receiver [9]. A periodic component induced by orbit eccentricity adds a relativistic effect due to the rate of change of eccentric anomaly. The correction for relativistic clock effects is formulated in the following manner.

$$\Delta t_{rel} = -2 \frac{\vec{r}_{sat} \cdot \vec{v}_{sat}}{c^2} \quad (4)$$

where, r_{sat} and v_{sat} are the LEO satellite position and velocity component vectors. The correction term Δt_{rel} has unit of time.

This secondary relativistic effect, known as the Shapiro [8, 15] signal propagation delay, is pertinent primarily in scenarios demanding high-precision positioning. Its overall impact on range is minimal, typically less

than 2 cm. Therefore, for the majority of applications, this effect can be safely disregarded. The Shapiro signal propagation delay introduces a correction to the geometric range based on general relativity principles. This correction is expressed by the following equation:

$$\Delta\rho_{rel} = \left(\frac{2\mu}{c^2}\right) ln\ ln\ \left(\frac{r^{sat}+r_{rcv}+r_{rcv}^{sat}}{r_{sat}+r_{rcv}-r_{rcv}^{sat}}\right) \quad (5)$$

8. Carrier ambiguity resolution: By referencing both code and carrier models discussed in Section C, the estimation of carrier ambiguity [16-18] for a specific time window can be achieved by subtracting ionospheric error-free code measurements from ionospheric error-free carrier measurements over a significant duration and subsequently averaging the differences. This process, encompassing epochs from (k) to $(k + n)$, where $(k + n)$ represents the last epoch before the satellite's measurement becomes unavailable, allows for the calculation of carrier ambiguity. The averaging effectively eliminates the Gaussian error component, leaving only the ambiguity value. The duration of the window is not fixed; rather, it relies on the future availability of measurements for the respective GPS satellite. Upon the reappearance of the satellite, carrier ambiguity is recalculated.

Table I: GPS Precise Positioning Service Typical UERE Budget type-1 [9]

Source	Error	1σ Error (m)
Space/ control	Broadcast Ephemeris	0.8
User	Residual Ionospheric Delay	0.1
	Residual Tropospheric Delay	0.2
	Receiver Noise and Resolution	0.1
	Multipath	0.2
System UERE	**Total**	**1.4**

Implementation of the measurement model

In a broad context, Global Navigation Satellite System (GNSS) range observations are typically characterized as one-way range observations. Specifically, range observations derived from code phase measurements are referred to as Pseudo-ranges. This concept is formally modeled as described in reference [7, 20].

$$\rho_i(t) = r_i(t) + c[\delta t_u(t) - \delta t_i(t-\tau)] + I(t) + T(t) + \epsilon_\rho(t) \quad (6)$$

Observations of range derived from carrier phase measurements are termed Carrier-ranges. This phenomenon is formally represented as outlined in reference [22].

$$\Phi_i(t) = r_i(t) + c[\delta t_u(t) - \delta t_i(t-\tau)] + I_\Phi(t) + T_\Phi(t) + \lambda N + \epsilon_\Phi(t) \qquad (7)$$

where i is the GNSS satellite index, ρ_i is the pseudo-range from the user's vehicle to the ith navigation satellite, Φ_i is the carrier range from the user's vehicle to the ith navigation satellite, $r_i(t)$ is the geometric distance from the user's vehicle to the ith navigation satellite, $\delta t_u(t)$ is the receiver clock offset, δt_i is the clock offset of the navigation satellite, τ is the signal transmission time, c is the velocity of light in space, $I(t)$ is the pseudo-range ionospheric error, $T(t)$ is the pseudo-range tropospheric error, $I_\Phi(t)$ is the carrier range ionospheric error, $T_\Phi(t)$ is the carrier range tropospheric error, λ is the wavelength of the carrier signal, N is integer ambiguity of the carrier signal, $\epsilon_\rho(t)$ is the random noise in the pseudo-range observation, $\epsilon_\Phi(t)$ is the random noise in the carrier-range observation

The measurement model is seamlessly integrated into the Kalman Filter's update process [21], wherein the state vector comprises three position components, three velocity components, and two clock components. The measurement equation guides the calculation of the Kalman Filter's observation matrix, achieved through partial differentiation with respect to the state variables. Additionally, the measurement error covariance matrix is selected in accordance with the presented error budget. The ensuing section presents the obtained results.

Result and discussion

The code or carrier residue represents the disparity between the a posteriori measurement, and the actual range measurements. These a posteriori range measurement are generated by passing the estimated state of the Extended Kalman Filter (EKF) through the code and carrier measurement functions specified in equations (6) and (7).

The genuine pseudo-range and carrier-range measurements are sourced from the RINEX observation file. The residue plots for both code and carrier are illustrated in Figures 1 and 2 throughout the entire day. The number of observations for both code and carrier of the GRACE satellite ranges from 5 to 10 during any given time instant. Distinct colors are employed to distinguish residues obtained from multiple pseudo-range observations.

Observations reveal that the majority of residues fall within the range of ±2m for code and ±0.5m for carrier. This observation suggests that both pseudo-range and carrier-range estimations derived from the EKF state are sufficiently accurate for position estimation. Notably, carrier-range estimation outperforms code, exhibiting much lower measurement residue, aligning with expectations.

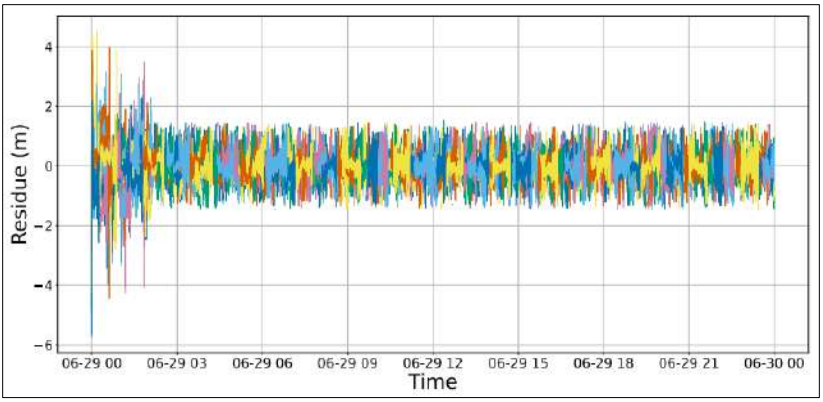


Figure 1: GRACE measurement residue plot for code

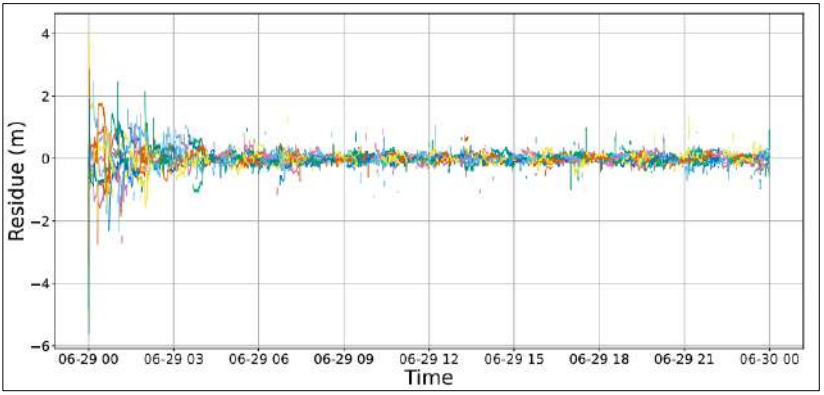
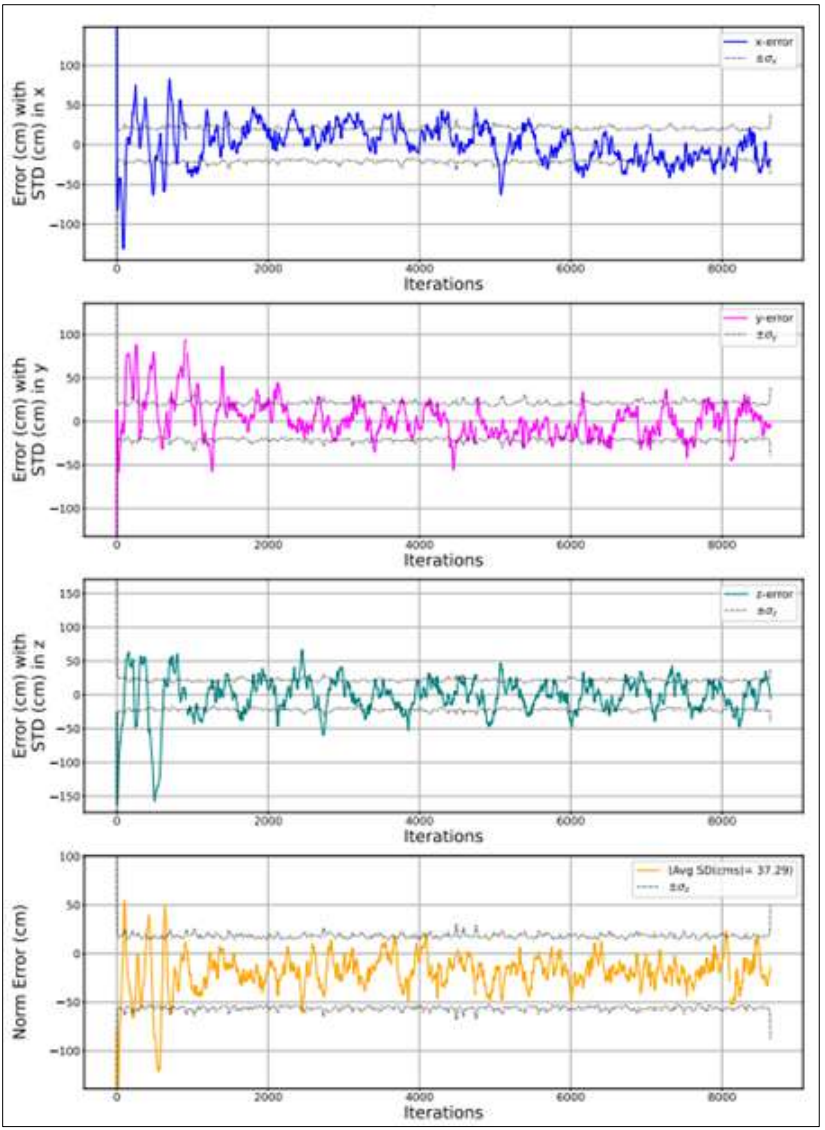


Figure 2: GRACE measurement residue plot for carrier


Error (cm) with STD (cm) in x
x-error
Iterations
Error (cm) with STD (cm) in y
y-error
Iterations
Error (cm) with STD (cm) in z
z-error
Iterations
Norm Error (cm)
Iterations

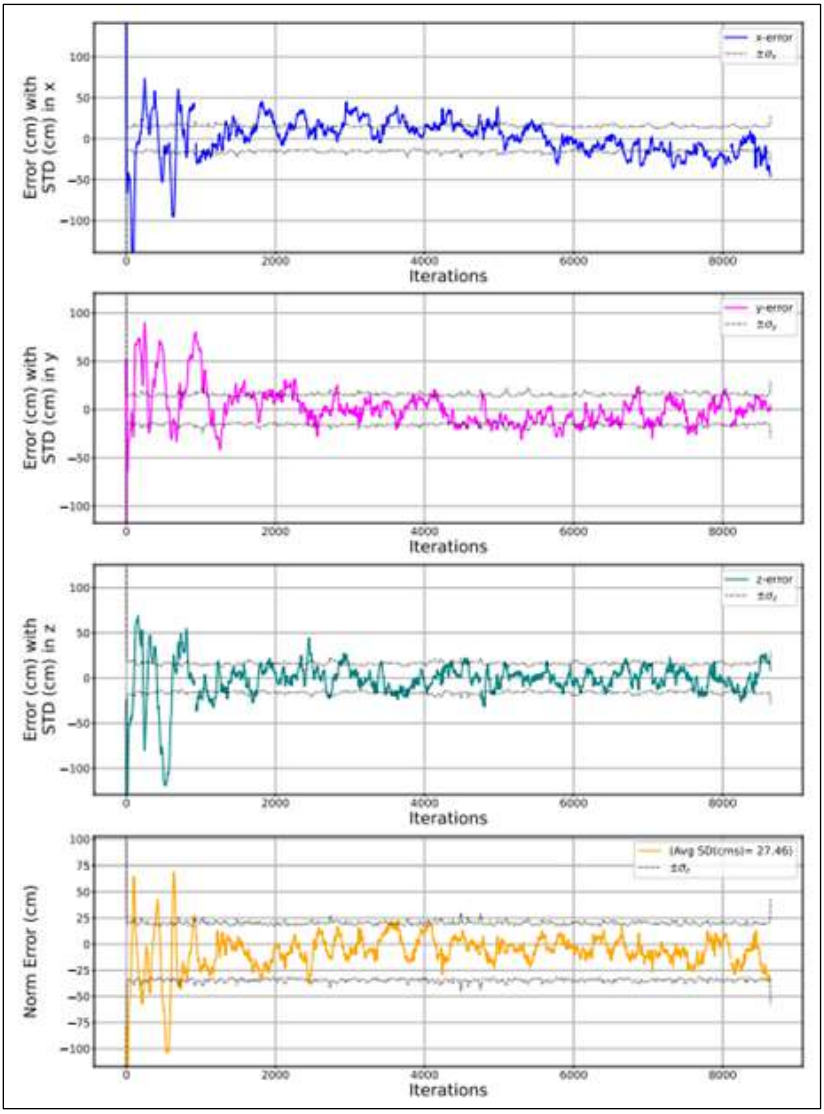


Figure 3: Position error plots for code (left) and carrier (right)

The calculation of position error involves determining the disparity between the estimated position r obtained from the Extended Kalman Filter (EKF) utilizing measurements and the actual position r derived from GRACE's kinematic orbit file. Figure 3 illustrate the position error in the estimation of the GRACE orbit using both code and carrier measurements.

To quantify the variability in position error, the standard deviation is computed for each axis. This standard deviation is derived as the square root of the first three diagonal elements of the updated (a posteriori) error covariance matrix.

A substantial portion of the data points falls within the ±σ uncertainty limit, and nearly all error data are encompassed within the ±3σ uncertainty range. The position error distribution along each axis exhibits a Gaussian pattern following the application of the Extended Kalman Filter (EKF), with a mean error close to zero. The utilization of carrier measurements results in a reduction of the mean range error compared to code measurements, aligning with the intended outcome.

Presently, available GPS receivers boast an average accuracy of approximately 10 meters. In the case of the GRACE mission, the position estimation accuracy is reported as less than -17.5 ± 13.6 cm for code measurements and less than -5.4 ± 11.0 cm for carrier measurements, respectively. Notably, these achieved accuracies closely rival those of commercial-grade Precise Point Positioning (PPP) GPS units.

Conclusion and future work

Raw GNSS observations are generated through simulation considering the effect of various factors like Satellite clock offset, Earth rotation during signal flight time, Relativistic Clock Correction, Ionospheric delay, etc. The predicted code and carrier phase range observations and the real navigation data are processed in an Extended Kalman Filter statistically to estimate the position and velocity and receiver clock bias of Low Earth Orbit in Earth-Centered Earth-Fixed frame. The Extended Kalman Filter is tuned to achieve good performance in terms of accuracy and a fixed interval smoothing algorithm is used to enhance the result in post-processing. The accuracy of the estimation process is approximately within -17.5 ± 13.6 cm for code -5.4 ± 11.0 cm for carrier measurement, which is better than the practical GPS receivers which provide a 10 m standard deviation.

The success of this study opens avenues for further exploration and improvement in the field of Precise Point Positioning by improving the measurement model. Future research could delve into the areas of using Multi-Constellation Integration [22], exploring advanced atmospheric models to correct ionospheric and tropospheric errors, integrating machine learning algorithms, etc.

References

1. Paziewski, Jacek. "Recent advances and perspectives for positioning and applications with smartphone GNSS observations." Measurement Science and Technology 31.9 (2020): 091001.
2. Ishikawa, Tadashi, *et al*. "History of on-board equipment improvement for GNSS-A observation with focus on observation frequency." Frontiers in Earth Science 8 (2020): 150.
3. Zhang, Baocheng, *et al*. "PPP-RTK functional models formulated with undifferenced and uncombined GNSS observations." Satellite Navigation 3.1 (2022): 3.
4. Nie, Zhixi, Fei Liu, and Yang Gao. "Real-time precise point positioning with a low-cost dual-frequency GNSS device." Gps Solutions 24 (2020): 1-11.
5. Wanninger, Lambert, and Anja Heßelbarth. "GNSS code and carrier phase observations of a Huawei P30 smartphone: Quality assessment and centimeter-accurate positioning." GPS Solutions 24.2 (2020): 64.
6. Magiera, Weronika, *et al*. "Accuracy of Code GNSS Receivers under Various Conditions." Remote Sensing 14.11 (2022): 2615.
7. S. K. Biswas, L. Qiao, and A. G. Dempster, "A Novel a Priori State Computation Strategy for the Unscented Kalman Filter to Improve Computational Efficiency," IEEE Transactions on Automatic Control, vol. 62, pp. 1852-1864, 2017.
8. Carlin, Luca, André Hauschild, and Oliver Montenbruck. "Precise point positioning with GPS and Galileo broadcast ephemerides." GPS Solutions 25.2 (2021): 77.
9. E. D. Kaplan and C. Hegarty, Eds., Understanding GPS: principles and applications, 2nd ed., ser. Artech House mobile communications series. Boston: Artech House, 2006
10. Sun, Peng, *et al*. "An investigation into real-time GPS/GLONASS single-frequency precise point positioning and its atmospheric mitigation strategies." Measurement Science and Technology 32.11 (2021): 115018.
11. Martens, Hilary R., *et al*. "Atmospheric pressure loading in GPS positions: Dependency on GPS processing methods and effect on assessment of seasonal deformation in the contiguous USA and Alaska." Journal of Geodesy 94 (2020): 1-22.

12. Xiong, Chao, *et al*. "On the occurrence of GPS signal amplitude degradation for receivers on board LEO satellites." Space Weather 18.2 (2020): e2019SW002398.
13. Gautier, Romain, *et al*. "Accurate measurement of the Sagnac effect for matter waves." Science Advances 8.23 (2022): eabn8009.
14. Wang, Dixing, *et al*. "Analysis of the J2 relativistic effect on the performance of on-board atomic clocks." GPS Solutions 27.3 (2023): 114.
15. Sośnica, Krzysztof, *et al*. "GPS, GLONASS, and Galileo orbit geometry variations caused by general relativity focusing on Galileo in eccentric orbits." GPS Solutions 26 (2022): 1-12.
16. El Rabbany, Ahmed. "The effect of physical correlations on the ambiguity resolution and accuracy estimation in GPS differential postioning." (2023).
17. Liu, Tianjun, *et al*. "Assessing GPS/Galileo real-time precise point positioning with ambiguity resolution based on phase biases from CNES." Advances in Space Research 66.4 (2020): 810-825.
18. Mohd Dzukhi, M. I., *et al*. "Reliability of the GPS carrier-phase fix solution under harsh condition." The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences 46 (2022): 223-227.
19. Leandro, Rodrigo. "Precise point positioning with GPS: A new approach for positioning, atmostpheric studies, and signal analysis." (2023).
20. Bramanto, Brian, Irwan Gumilar, and Irma AN Kuswanti. "Assessment of GNSS observations and positioning performance from non-flagship android smartphones." Journal of Applied Geodesy 0 (2023).
21. Elmezayen, Abdelsatar, and Ahmed El-Rabbany. "Real-time GNSS precise point positioning using improved robust adaptive Kalman filter." Survey Review 53.381 (2021): 528-542.
22. An, Xiangdong, Xiaolin Meng, and Weiping Jiang. "Multi-constellation GNSS precise point positioning with multi-frequency raw observations and dual-frequency observations of ionospheric-free linear combination." Satellite navigation 1.1 (2020): 1-13.

Chapter - 4

Enhanced Performance in Multiple Object Tracking: Kalman Filter Integration with Hungarian Algorithm and RMS Index Optimization

Authors

Sk Babul Akhtar
Swami Vivekananda University, Barrackpore, North 24 Parganas, Kolkata, West Bengal, India

Pradipta Maiti
Swami Vivekananda University, Barrackpore, North 24 Parganas, Kolkata, West Bengal, India

Chapter - 4

Enhanced Performance in Multiple Object Tracking: Kalman Filter Integration with Hungarian Algorithm and RMS Index Optimization

Sk Babul Akhtar and Pradipta Maiti

Abstract

Object tracking plays a vital role in various computer vision and robotics applications, such as surveillance, autonomous navigation, and human-computer interaction. The Kalman Filter (KF) stands out as a prominent tool for object tracking due to its efficiency and adaptability in handling noisy measurements and dynamic environments. This study utilizes the Kalman Filter to track both simulated objects and real-life video tracking in a 2D context. By employing the Root Mean Square (RMS) index, we fine-tune the Kalman filter for optimal results based on specific parameter requirements. The paper begins with an in-depth exploration of the Kalman Filter, covering its mathematical foundations and core principles. It then introduces the concept of the RMS index and presents an algorithm that integrates both elements into a unified approach. Additionally, the research employs the Hungarian algorithm to track multiple objects in video footage. A thorough comparative analysis of results with different tuning parameters illustrates the superior performance of the customized settings in specific scenarios. In conclusion, our findings confirm that a properly tuned Kalman Filter operates with low-power computation and delivers exceptional efficiency compared to alternative methodologies.

Keywords: Kalman filter, object tracking, RMS index, 2D tracking, hungarian algorithm.

Introduction

Object Tracking plays a crucial role in various applications within the realm of computer vision. Object Tracking algorithms have garnered significant attention with the advancement of powerful computing systems,

the widespread availability of affordable high-resolution video cameras, and the growing demand for automated video analysis [1, 2]. Among the myriad tracking algorithms available, the Kalman Filter (KF) has consistently emerged as a powerful tool due to its efficiency and adaptability in handling noisy measurements and dynamic environments [3-5]. This section aims to provide insights into the latest developments in Kalman Filter-based object tracking, highlighting its advantages over alternative methods, and offering a selection of recent references that reflect the current state of the field. Particle filters are a popular alternative to the KF for non-linear and non-Gaussian tracking problems. While particle filters can handle non-linearities and multi-modal distributions better than Kalman Filters, they suffer from high computational complexity [6]. Mean-shift tracking is particularly effective for object tracking in cluttered scenes and handles appearance changes well. However, mean-shift may struggle with scale variations and abrupt object motion [7]. Template matching is simple and computationally efficient, but can be sensitive to variations in object appearance, illumination changes, and occlusions.

In summary, while there are various object tracking methods available, the Kalman Filter remains a popular choice for object tracking due to its computational efficiency, simplicity, and noise-handling capabilities. While other methods like particle filters and mean-shift tracking excel in specific scenarios, the Kalman Filter offers a balanced approach suitable for a wide range of tracking applications [8-11].

Theoretical background

In this section, the paper discusses the underlying theory of object tracking using the Kalman Filter and various other algorithm components related to the overall scope. The section concisely overviews the Kalman Filter theory [6] and the RMS index [15]. It's worth noting that more in-depth explanations of these topics can be found in other papers [12-16], as they exceed the scope of this one. The section also includes the presentation of equations for tracking an object in two dimensions and the introduction of the blob detection algorithm, which is used to detect simulated objects or blobs in a uniform background. Additionally, in order to achieve real video tracking, the paper makes use of the Hungarian algorithm [17] in tandem with the foundational KF theory, as discussed earlier. It is important to highlight that real-time video tracking leverages pre-trained weights from a Deep

Neural Network (DNN) [18, 19] implemented using the OpenCV library in Python. These pre-trained weights aid in object detection, generating the measurements essential for the functioning of the Kalman Filter system.

Kalman filter

The Kalman filter works in two main parts: prediction and update. The state prediction is given by Equation (1) and covariance prediction is given by Equation (2):

$$\hat{x}_k^- = \varphi \hat{x}_{k-1}^+ + B u_{k-1} \quad (1)$$

$$\hat{P}_k^- = \varphi \hat{P}_{k-1}^+ \varphi^T + Q \quad (2)$$

where $\hat{x}_k^-$ is the predicted state at time k, φ is the state transition matrix at time k, B is the input matrix at time k, u_{k-1} is the input at time k, $\hat{P}_k^-$ is the predicted covariance matrix at time k, and Q is the process noise covariance matrix at time k.

For the KF update step, the relevant equations, namely, measurement residual, residual covariance, Kalman gain, state update, and covariance update are given from Equation (3) to Equation (7) respectively:

$$y_k = z_k - H\hat{x}_k^- \quad (3)$$

$$S_k = H\hat{P}_k^- H^T + R_k \quad (4)$$

$$K_k = \hat{P}_k^- H^T {S_k}^{-1} \quad (5)$$

$$\hat{x}_k^+ = \hat{x}_k^- + K_k y_k \quad (6)$$

$$P_k^+ = (I - K_k H) P_k^- (I - K_k H)^T + K_k R_k \quad (7)$$

where z_k is the measurement at time k, H is the measurement matrix at time k, R_k is the measurement noise covariance matrix at time k, y_k is the measurement residual at time k, S_k is the residual covariance at time k, K_k is the Kalman gain at time k, $\hat{x}_k^+$ is the updated state at time k, and P_k^+ is the updated covariance matrix at time k.

Considering a 2D tracking system, the following KF equations are derived from the standard KF equations. The state vector can be extended to the previous state at time, $(t-1)$ using Equation (8). Here, x_t and y_t are the concerned variables denoting the object position and dt is the sampling interval.

$$X_t = [x_t\ y_t\ \dot{x}_t\ \dot{y}_t\ ] =$$
$$[1\ 0\ dt\ 0\ 0\ 1\ 0\ dt\ 0\ 0\ 1\ 0\ 0\ 0\ 0\ 1\ ][x_{t-1}\ y_{t-1}\ \dot{x}_{t-1}\ \dot{y}_{t-1}\ ] +$$

$$\left[\frac{1}{2}dt^2\ 0\ 0\ \frac{1}{2}dt^2\ dt\ 0\ 0\ dt\ \right][\ddot{x}_{t-1}\ \ddot{y}_{t-1}\ ] \quad (8)$$

From Equation (8), it is clear that:

$$\varphi = [1\ 0\ dt\ 0\ 0\ 1\ 0\ dt\ 0\ 0\ 1\ 0\ 0\ 0\ 0\ 1\ ]\ and\ B = \left[\frac{1}{2}dt^2\ 0\ 0\ \frac{1}{2}dt^2\ dt\ 0\ 0\ dt\ \right] \quad (9)$$

Since, only using only the position parameters are recorded for measurement, H can be stated as:

$$H = [1\ 0\ 0\ 0\ 0\ 1\ 0\ 0\ ] \quad (10)$$

Also, the standard process noise covariance, Q and measurement noise covariance, R for 2D tracking are given by Equation (12) and (11). ${\sigma_{\ddot{x}}}^2$ is the parameter that is needed to be tuned to get the best KF outcome. ${\sigma_x}^2$ and ${\sigma_y}^2$ are the variances for measurements in x and y coordinates.

$$R = \left[{\sigma_x}^2\ 0\ 0\ {\sigma_y}^2\ \right] \quad (11)$$

$$Q = \left[\frac{dt^4}{4}\ 0\ \frac{dt^3}{2}\ 0\ 0\ \frac{dt^3}{2}\ 0\ \frac{dt^4}{4}\ 0\ \frac{dt^3}{2}\ 0\ dt^2\ 0\ \frac{dt^3}{2}\ 0\ dt^2\ \right]{\sigma_{\ddot{x}}}^2 \quad (12)$$

RMS index

To calculate the RMS index for a POM system [16], the Kalman filter estimates the true state of the system based on the available position measurements, and then compares the estimated state to the true state over a specified time period. The RMS index is to be calculated for varying values of Q prior to running the KF. The Q for which RMS index provides the least value is then used to run the KF tracking system. The detailed algorithm to calculate RMS index can be found in [15, 16].

Assuming that the process noise in x and y coordinates are similar, Q for one-dimensional (1D) tracking can be extracted from Equation (), and can be expressed as $\begin{bmatrix} \frac{dt^4}{4} & \frac{dt^3}{2} \\ \frac{dt^3}{2} & dt^2 \end{bmatrix} \times \sigma_q^2$ or $[a\ b\ b\ c\ ]$. Now, considering the above expression, RMS index can be calculated using the following expression:

$$\mu_p = \frac{{a_D}^2}{\beta^2} + \frac{2\alpha^2 + 2\beta + \alpha\beta}{\alpha(4 - 2\alpha - \beta)} \quad (13)$$

where,

$$\beta = \frac{C+\sqrt{C(16+4A-4B+C)}}{4} - \sqrt{\frac{C^2(16+4A-4B+C)}{8\sqrt{C(16+4A-4B+C)}} + \frac{C(2A-2B+C)}{8}} \quad (14)$$

$$\alpha = 1 - \frac{\beta^2}{C} \quad (15)$$

and, A= $\frac{a}{B_x}$, B = $\frac{bdt}{B_x}$, C = $\frac{cdt^2}{B_x}$, ${a_D}^2 = \frac{{a_c}^2 dt^4}{B_x}$

Here, a_c is assumed to be the constant acceleration of the moving object, which is assumed to be estimated. B_x is the measurement noise variance, i.e., ${\sigma_x}^2$ or ${\sigma_y}^2$, with the assumption that both the measurement noise variances are equal. If in case, either of measurement noise variance or process noise variance for x and y coordinates differ, then two separate RMS indices needs to be calculated. The choice to assume equal noise variances for simplicity in practical object tracking is made with this context in mind.

Blob detection and hungarian algorithm

A basic blob detecting algorithm is used to detect moving circles in a video sequence, on which the 2D tracking test is performed. This method finally outputs the centroid of the object, which is used in the KF measurement model. The first step is to convert the image from color to grayscale which simplifies the image by removing color information. Then, an edge detection algorithm, such as the Canny edge detector [20] or the Sobel operator [21], is applied to the grayscale image which highlight areas of significant intensity variation, denoting edges or boundaries of objects. After converting the image to binary, contour detection algorithms, such as the OpenCV function findContours(), are applied. Finally, detected contours are used to identify blobs or objects.

When detecting multiple objects in a frame, the image processing algorithm treats each frame as an isolated task, providing centroids in the order of detection, thus failing to track multiple objects. The challenge is to accurately assign IDs to objects detected in the next frame. Considering objects detected in two consecutive frames, with edges indicating potential matches based on Euclidean distances, the objective is to find the minimum-weight matching to correctly ID objects from past frame to current. To address this issue efficiently, the Hungarian algorithm [17, 22] is employed, processing the bipartite graph represented as an adjacency matrix.

Proposed model

The diagram in Fig. 1 shows the detailed block diagram to track

multiple objects in a video sequence.

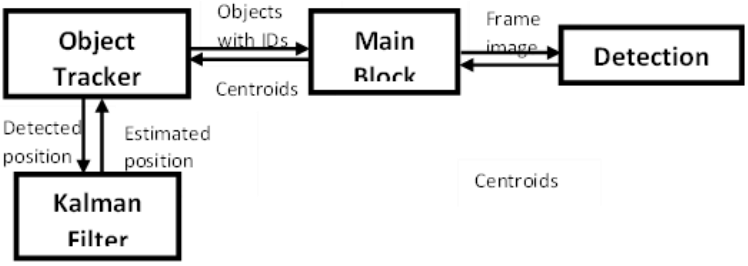


Figure 1: Block diagram for object tracking

The main block is responsible to sending image frames and generating output video. Object Tracker keeps tabs on objects by using the Hungarian algorithm mentioned in Section 2.3. Meanwhile, the Kalman filter predicts and figures out where each object is in the next frames, using data and a model. This filter's accuracy is adjusted using the RMS index, calculated using the method explained in Section 2.2. Finally, the detection block, as described in Section 2.3, identifies the center points of objects for tracking by using edge detection.

Results and discussion

Simulated Object 2D tracking

Fig. 2 presents a visual representation of a frame, displaying both the actual measured position and the predicted estimated position. In Fig. 3, there's a plot that charts the coordinates of an object moving in two dimensions, comparing the estimated and measured values. Fig. 4 provides a graphical depiction of how far the object's position deviates from the true coordinates, showing the absolute deviation.

Figure 2: Graphical representation of 2D tracking

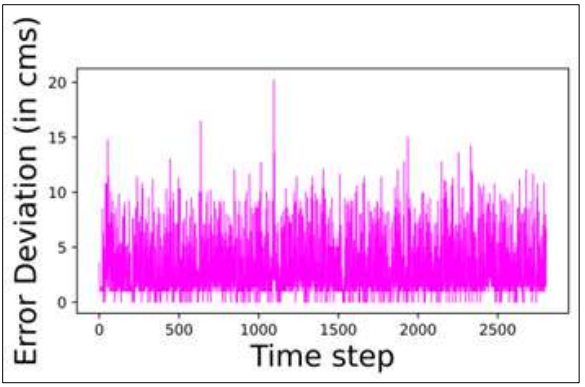


Figure 3: Estimated path vs. true path

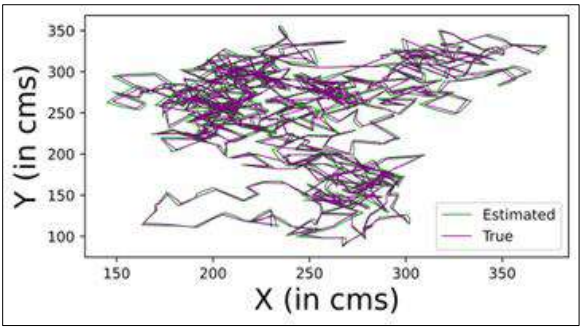


Figure 4: Error deviation from true path

Table 1 shows the various errors for different Q values, thus concluding that the tuned Q for $p = 1$, which is derived from the discussed algorithm provides the best result. The term p is used in context of the relation ($Q = Q_{nom} \times 10^{p}$) using which Q is varied, where Q_{nom} is the nominal Q.

Table 1: Performance comparison for different Q values

p value	RMS Index	Abs. Mean Error (cm)
-1	11.6	7.73
0 (nominal)	1.49	4.90
1 (tuned)	0.79	2.89
2	1.67	5.02
3	4.62	7.10

Real video tracking

Fig. 5 shows the simulation for tracking in video sequence. Here the same algorithm is used as discussed in Section 2.4 to track pedestrians in a video sequence. The extra layer that is added in order to do so uses the DNN based detection technique using pre-trained weights, which is integrated within the KF measurement model.

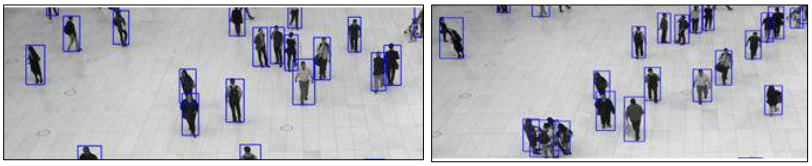


Figure 5: Real video tracking

Conclusion and future work

In our exploration of Object Tracking using the Kalman filter, we have conducted a comprehensive analysis. Our findings from the simulation results are significant. It is evident that the Kalman filter performs exceptionally well when certain conditions are met: the system model is linear, the noise distribution is assumed to be Gaussian, and the filter is finely tuned. We have also addressed to the challenge of Multiple Object Tracking, by efficiently applying the Hungarian algorithm. Tuning the noise parameters of the Kalman filter can be a complex endeavor, and in this thesis, we have introduced an approach to automatically adjust the process noise covariance matrix, Q. Our results in Section 3.1 clearly demonstrate that with proper tuning, a poorly-performing system can experience a substantial enhancement in its overall performance.

For future research, this method can be further explored, tested, and adapted to non-linear and more intricate models to assess its versatility. Additionally, the algorithm's suitability for different variants of the Kalman filter, such as the Unscented Kalman Filter (UKF) and Extended Kalman Filter (EKF), could also be investigated.

References

1. Cedras, C., Shah, M.: Motion-based recognition a survey. In: Image and vision computing, vol. 13(2),pp. 129-155. (1995).
2. Yilmaz, A., Javed, O. and Shah, M.: Object tracking: A survey. In: Acm computing surveys (CSUR), vol. 38(4), pp.13-es, (2006).

3. Basso, G.F., De Amorim, T.G.S., Brito, A.V. and Nascimento, T.P.: Kalman filter with dynamical setting of optimal process noise covariance. In: IEEE Access, vol. 5, pp.8385-8393, (2017).

4. Mirunalini, P., Jaisakthi, S.M. and Sujana, R.: Tracking of object in occluded and non-occluded environment using SIFT and Kalman filter. In: TENCON 2017-2017 IEEE Region 10 Conference (pp. 1290-1295), (2017).

5. Jiao, J. and Wang, H.: Traffic behavior recognition from traffic videos under occlusion condition: a Kalman filter approach. In: Transportation research record, 2676(7), pp.55-65, (2022).

6. Bukey, C.M., Kulkarni, S.V. and Chavan, R.A.: Multi-object tracking using Kalman filter and particle filter. In: 2017 IEEE International Conference on Power, Control, Signals and Instrumentation Engineering (ICPCSI) (pp. 1688-1692), (2017).

7. Kumar, S., Raja, R. and Gandham, A.: Tracking an object using traditional MS (Mean Shift) and CBWH MS (Mean Shift) algorithm with Kalman filter. In: Applications of Machine Learning, pp.47-65, (2020).

8. Jondhale, S.R. and Deshpande, R.S.: Kalman filtering framework-based real time target tracking in wireless sensor networks using generalized regression neural networks. IEEE Sensors Journal, vol. 19(1), pp.224-233, (2018).

9. Xu, S. and Chang, A.: Robust object tracking using Kalman filters with dynamic covariance. Cornell University, pp.1-5, (2014).

10. Mahalingam, T. and Subramoniam, M.: Optimal object detection and tracking in occluded video using DNN and gravitational search algorithm.In: Soft Computing, vol. 24, pp.18301-18320, (2020).

11. Barrau, A. and Bonnabel, S.: The invariant extended Kalman filter as a stable observer. In: IEEE Transactions on Automatic Control, vol. 62(4), pp.1797-1812, (2016).

12. Yang, F., Chen, H., Li, J., Li, F., Wang, L. and Yan, X.: Single shot multibox detector with kalman filter for online pedestrian detection in video. In: IEEE Access, vol. 7, pp.15478-15488, (2019).

13. Chen, Y., Zhao, D. and Li, H.: Deep Kalman filter with optical flow for multiple object tracking. In: 2019 IEEE international conference on systems, man and cybernetics (SMC) (pp. 3036-3041), (2019).

14. Fauzi, N.I.H., Musa, Z. and Hujainah, F.: Feature-Based Object Detection and Tracking: A Systematic Literature Review. International Journal of Image and Graphics, p.2450037, (2023).
15. Saho, K. and Masugi, M.: Automatic Parameter Setting Method for an Accurate Kalman Filter Tracker Using an Analytical Steady-State Performance Index. In: IEEE Access, vol. 3, pp. 1919-1930, (2015).
16. Saho, K.: Kalman Filter for Moving Object Tracking: Performance Analysis and Filter Design, 10.5772/intechopen.71731, (2018).
17. Hamuda, E., Mc Ginley, B., Glavin, M. and Jones, E.: Improved image processing-based crop detection using Kalman filtering and the Hungarian algorithm. In: Computers and electronics in agriculture, vol. 148, pp.37-44, (2018).
18. Erol, B. A., Majumdar, A., Lwowski, J., Benavidez, P., Rad, P., and Jamshidi, M.: Improved deep neural network object tracking system for applications in home robotics. In: Computational Intelligence for Pattern Recognition, pp. 369-395, (2018).
19. Szegedy, C., Toshev, A., & Erhan, D.: Deep neural networks for object detection. Advances in neural information processing systems, vol. 26, (2018).
20. Sekehravani, E. A., Babulak, E., and Masoodi, M.: Implementing canny edge detection algorithm for noisy image. Bulletin of Electrical Engineering and Informatics, vol. 9(4), pp. 1404-1410, (2020).
21. Han, L., Tian, Y., & Qi, Q.: Research on edge detection algorithm based on improved sobel operator. In: MATEC Web of Conferences (Vol. 309, p. 03031). EDP Sciences, (2020).

Chapter - 5

Computational Study of FeAsn (n=1-4) Clusters Invoking DFT Based Descriptors

Authors

Shayeri Das
Department of Electrical Engineering, St. Mary's Technical Campus Kolkata, Saibona, West Bengal, India
Department of Mechatronics Engineering, Manipal University Jaipur, Dehmi Kalan, Rajasthan, India

Tomal Suvro Sannyashi
Department of Electronics and Communication Engineering, Swami Vivekananda University, Barrackpore, West Bengal, India

Dr. Debasis Mondal
Department of Electronics and Communication Engineering, Swami Vivekananda University, Barrackpore, West Bengal, India

Prabhat Ranjan
Department of Mechatronics Engineering, Manipal University Jaipur, Dehmi Kalan, Rajasthan, India

Tanmoy Chakraborty
Department of Chemistry and Biochemistry, School of Basic Sciences and Research, Sharda University, Greater Noida, Uttar Pradesh, India

Chapter - 5

Computational Study of FeAsn (n=1-4) Clusters Invoking DFT Based Descriptors

Shayeri Das, Tomal Suvro Sannyashi, Dr. Debasis Mondal, Prabhat Ranjan and Tanmoy Chakraborty

Abstract

A study of FeAsn (n=1-4) is performed invoking Density Functional Theory (DFT) methodology due to its application in superconductors. Geometry optimization is implemented by using exchange correlation B3LYP with basis set LANL2DZ. The computed bond lengths and global descriptors like molecular hardness, softness, electronegativity, electrophilicity index of the most stable configurations of the clusters are also in line with their experimental values. The results show that all the clusters, due to the substantial highest occupied molecular orbital (HOMO) - lowest unoccupied molecular orbital (LUMO) gap, have good prospects in the superconductivity applications.

Keywords: Density functional theory, As-Fe; HOMO-LUMO energy gap, superconductors.

Introduction

Superconductivity has attracted a lot of interest in recent years especially in the domain of solid-state physics [1]. The main aspects of interest in the field of superconductivity are study of high critical temperature materials and unveiling the secret of superconducting electron pairs [1]. The discovery of superconductivity in rare earth oxypnictides compounds have inspired researchers. Johnson *et al.* [2] reported the superconductivity of quaternary ZrCuSiAs compound, which paved the way for all further research. Compounds with rich chemical compositions were synthesized, that exhibited a range of electronic properties like ferromagnetism, antiferromagnetism, semi conductivity, superconductivity etc. [3-12].

Kamihara *et al*. [13] were the first to discover superconductivity in LaFePO at relatively low critical temperature. Initially experimental research was performed based of clusters containing elements exhibiting superconductivity [14-17]. Arsenic based superconductors have been reported with doping of other elements, which enhance their properties [18-23]. Katayama *et al*. [18] reported superconductivity of Ca1-xLaxFeAs2. Single-crystal X-ray diffraction analysis revealed that the crystallization occurred within the clusters in zigzag chains, forming monoclinic structures. Yakita *et al*. [19] performed a study on CaFeAs2 and PrFeAs2 where they studied the magnetization along with resistivity measurements of the clusters. The clusters unveiled superconductivity mutually in magnetization and resistivity measurements. Kudo *et al*. [20] reported a huge variation in critical temperature with phosphorus and antimony doping of Ca1−xLaxFeAs2, tin doping the same cluster, the crystals were fabricated experimentally, and X-ray diffraction helped in analysis of the crystals. Kudo *et al*. [21] studied the increase in critical temperature of 112 phases of Ca1−xRExFeAs2 and observed the variation of critical temperature is greater for La doped clusters than for Pr and Nd doped samples. Bert *et al*. [22] reported the structural properties of GaAs, in which arenic cluster size is established attributing to superconducting phase.

Similarly, iron doped superconductors, which are also considered high temperature superconductors, were first reported by Gordon *et al*. [24]. These new materials offered scholars a new perspective of these superconductors, yet also furnished variable innovative concepts for scrutinizing the superconductivity contrivances, comprising antiferromagnetic variations [25] designated in form of a spin density wave [25], pseudo gap [26], and three-band theory [27]. Bardeen-Cooper-Schrieffer theory or BCS Theory was the first theory of superconductivity describing superconductivity as a microscopic effect caused due to condensation of Cooper pairs [28]. Iron atoms have unpaired electrons in the outer orbitals, yet are still aligned with one another [29]. Iron is a good conductor, but it is not clear how they could become zero-resistance superconductors at high temperatures without the strong interactions that produce a correlated insulating state in the copper-based materials [29]. As a result, “Fe-based superconductors” cannot be clarified by Bardeen-Cooper-Schrieffer theory or BCS theory, and are consequently termed as unconventional superconductors [30, 31].

Experimental facts specify that the Fe-As-based superconductors have a much higher critical temperature than others have and, furthermore, only

certain type Fe-As-based superconductors exhibited a high critical temperature above 40 K [21]. Theoretical studies have shown that some arsenic based superconductors exhibit similar electronic band structures [32]. The focus of our study is thus based on Fe-As based clusters, the properties of their optimized structures and their prospects. This report has been presented in such a way that Section 2 provides a brief account of the computational methodology. Additionally, the subsequent sections provides the computational results along with the corresponding discussions. The entirety of the study has been briefed in section 4.

Computational details

Numerous fields of study like alloy based clusters, fluid mechanics, molecular and nuclear physics, life sciences etc. has adopted this technique [33-38]. DFT methods is successfully used in various cases [39-42]. In the past few years, our group has executed various analysis on metallic clusters [43-49] based on DFT.

In this report a theoretical investigation based on different nanoclusters, FeAsn (n=1-4) has been conducted. Optimization of the clusters have been instigated by means of Gaussian 16 [50] using DFT framework. The hybrid functional parameter (B3LYP) aided by basis set LANL2DZ has been opted and intended for orientational optimization.

Ionization Energy (I) in addition to Electron Affinity (A) of all the clusters are estimated via the subsequent equations [51]:

$$I = - \varepsilon HOMO \quad \text{(i)}$$

$$A = - \varepsilon LUMO \quad \text{(ii)}$$

By means of values of 'I' and 'A', which are the conceptual descriptors namely molecular hardness (η), molecular softness (S), electro-negativity (χ) and electro-philicity index (ω) are worked out. The equations based on which computation have been performed have been presented in equation (iii) to equation (vi) -

$$\chi = -\mu = \frac{I + A}{2} \quad \text{(iii)}$$

Here, μ implies the chemical potential of the cluster.

$$\eta = \frac{I - A}{2} \quad \text{(iv)}$$

$$S = \frac{1}{2\eta} \tag{v}$$

$$\omega = \frac{\mu^2}{2\eta} \tag{vi}$$

Results and discussion

Equilibrium geometry

The ground state or least energy structures of FeAsn (n=1-4) has been presented in figure 1, along with the point groups and spin multiplicity. The figure also shows the bond lengths between all the atoms in every structure. Bond length is elucidated as the shortest path between the centres of atoms that are bonded. The dimension of the bond is confirmed using the aggregate value of bonded electrons. The increase in the magnitude of bond order, leads to the resilient attraction between the two atoms and results in smaller magnitude of bond length. Over all, the measurement of the bond amid two atoms is almost equivalent to the summation of the covalent radii of the concerned atoms. Bond length is represented in Angstrom (Å).

FeAs displays ground state configuration is a linear structure presented in 1-a with As atom linked in a line with Fe atom. The orientation has the symmetry group C∞V and doublet spin multiplicity. The bond length of As-Fe is 2.370Å.

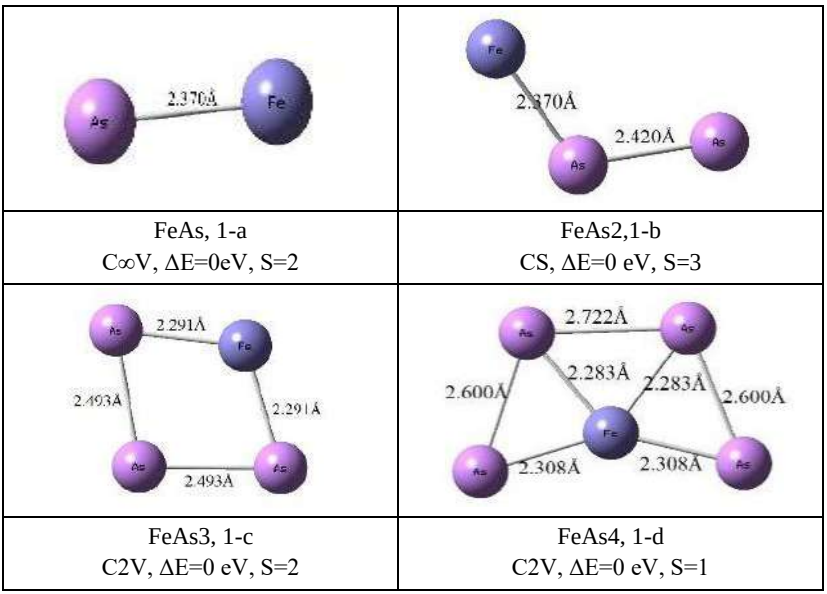
FeAs2 has the ground state structure, presented in 1-b, which has Fe placed at one of the ends. The structure comprises of point group of CS with triplet spin multiplicity. The bond length of As-As is 2.420 Å and that of As-Fe is 2.370 Å.

FeAs3 in 1-c represents a quadrilateral structure with the atoms connected to each other, exhibiting point group C2v and doublet spin multiplicity. The As-As bonds connected directly has bond length 2.493Å. Similarly, the As-Fe atoms, which are directly placed at vertices, possess a bond length of 2.291Å.

The structure possessing lowest energy for FeAs4 is presented in 1-d, which is the most stable structure, point group C2v and singlet spin multiplicity. The structure resembles a trapezium like shape. The As-As bond length at the non-parallel sides are both 2.600 Å. The As-As bond length at top side is 2.722 Å. The Fe atom in placed in the centre of the base

and the four As atoms are linked with it. The As atoms on the same side create As-Fe bonds of the length 2.308 Å and that on the opposite arm have bond lengths 2.283 Å.

The reported experimental bond length of As-As atoms is 2.414 Å-2.647 Å [52]. The computed data obtained for As-As bond length is in range with the experimentally reported value as the shortest bond length is approximately equal to the experimental value with a mere variation. The shortest bond length of As-As bond is 2.423 Å which is exhibited by FeAs2 and the longest bond length for the pair is 2.722Å which is also possessed by FeAs4.



FeAs, 1-a C∞V, ΔE=0eV, S=2	FeAs2,1-b CS, ΔE=0 eV, S=3
FeAs3, 1-c C2V, ΔE=0 eV, S=2	FeAs4, 1-d C2V, ΔE=0 eV, S=1

Figure 1: Lowest energy structures of FeAsn (n=1-4)

Electronic properties and DFT based descriptors

Using electronic structure theory, the computational study of arsenic-iron cluster is performed. The global descriptors that are established on density functional theory like molecular hardness, molecular softness, electronegativity and electrophilicity are evaluated using equations (iii) to (vi).

It has been stated that donor and acceptor orbitals have an imperative part in transference of charge and bond creation at the time of the setting up of donor-acceptor system [53-59]. The energy alteration amongst orbitals

HOMO and LUMO is an essential element to figure out the electronic characteristics like chemical hardness, softness, and others, values of which are determined by the energy gap of HOMO and LUMO [60, 61]. Kinetic stability of a cluster declines with the diminution of HOMO-LUMO gap. It states the minimal energy requirement of an electron to transfer from occupied orbital to unoccupied orbital [58]. In this study, FeAs2 exhibits the uppermost HOMO-LUMO gap and FeAs4, the least value. In the range obtained, all the clusters are suitable for application as superconductors.

In order to comprehend the dynamics of the cluster formed, molecular hardness is a vital factor [62]. It is the natural tendency of every molecule to configure themselves with maximum hardness [63]. Molecular hardness is an electronic factor of clusters that depict the constancy of the molecular system. Movement of a molecule from stable to unstable state reduces its hardness value and vice versa [55]. According to computed date, FeAs2 exhibits the highest value and FeAs4, the least, the range varying from 1.322eV to 1.659eV. The cluster with the peak HOMO-LUMO gap also possesses the highest magnitude of hardness and vice versa. The variation of hardness with HOMO-LUMO gap is presented in figure 2. The figure also represents the change in hardness with variation of cluster size with increase in number of arsenic molecules.

The molecular softness shares an inversely relation with HOMO-LUMO gap of the cluster. Correspondingly the cluster with the maximum hardness gap exhibits minutest softness and vice versa. The molecular softness ranges from 0.301eV - 0.378eV, with FeAs4 being highest and FeAs2 presenting with the least value.

Another crucial factor to comprehend the transfer of charge from donor to acceptor is electronegativity [64, 65]. In the obtained data, the electronegativity fluctuates from 4.343eV to 5.116 eV. Maximum value is for FeAs4 and the minimum for FeAs. In case of electronegativity all, the computed values are close in magnitude with no huge variations in value.

The electrophilicity index is the constraint for computation of the expanse of energy reduction due to additional flow of electrons for the duration of orbital interaction that is predisposed with both ionization energy and electron affinity [66]. The cluster possessing the lowest HOMO-LUMO gap exhibits the largest electrophilicity index. The computed data exhibits that electrophilicity index rises with the escalation in number of As atoms in the clusters.

Most of the structures except FeAs4 exhibit high dipole moment indicating their ionic nature of bond. Dipole moment is related to the structural symmetry of a cluster, among the clusters FeAs2 presents with the highest value, which also possess the highest HOMO-LUMO gap.

Table 1: DFT based global descriptors of FeAsn (n=1-4) clusters

Species	HOMO-LUMO Gap (in eV)	Hardness (eV)	Softness (eV)	Electronegativity (in eV)	Electrophilicity Index (in eV)	Dipole Moment (in Debye)
FeAs	3.138	1.569	0.319	4.343	6.010	3.119
$FeAs_2$	3.318	1.659	0.301	4.525	6.172	3.784
$FeAs_3$	2.866	1.433	0.349	4.951	8.552	2.470
$FeAs_4$	2.645	1.322	0.378	5.116	9.897	0.062

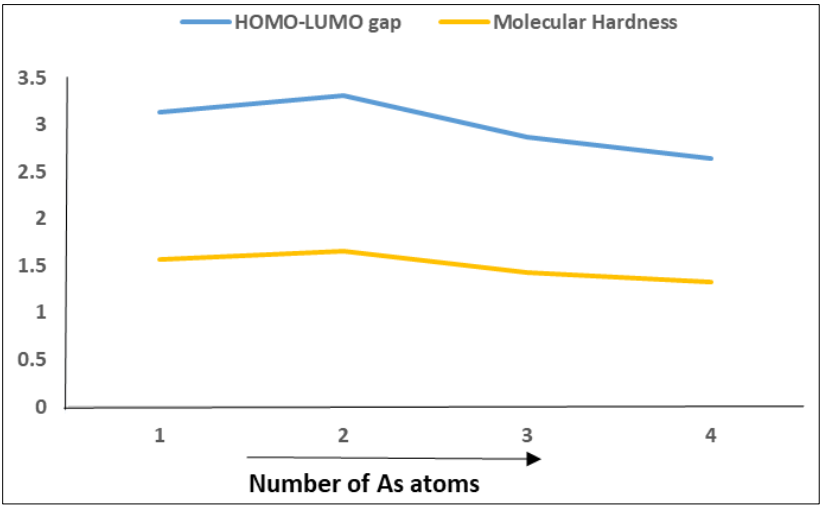


Figure 2: Relation of HOMO-LUMO gap with molecular hardness

Conclusion

A computational study of arsenic doped iron clusters have been performed. DFT based global descriptors have been computed based on the clusters varying 1 to 4 arsenic atoms with a single Fe atom, possessing the most stable configuration. The obtained bond length of the most stable state configuration of the clusters validate their reported experimental values. The computed data present with HOMO-LUMO gap, which is validated by reported values of metallic clusters. The nanoalloy clusters present with

energy gap, which is suitable for superconductor applications. The minute variation of molecular hardness and softness indicates slow progression in the change in properties of the clusters. The ionic bond formed within the cluster is clearly depicted by all the clusters except FeAs4. The obtained results have opened up avenues of new research in this field.

References

1. Ren ZA, Zhao Z X. Research and Prospects of Iron-Based Superconductors. Adv. Mater. 2009; 21: 4584-4592. DOI: https://doi.org/10.1002/adma.200901049
2. Johnson V, Jeitschko W. ZrCuSiAs: A “filled” PbFCl type. J. Solid State Chem. 1974; 11: 161-166. DOI: https://doi.org/10.1016/0022-4596(74)90111-X
3. Kaczorowski D, Albering JH, Noel H, Jeitschko W. Crystal structure and complex magnetic behaviour of a novel uranium oxyphosphide UCuPO. J. Alloys Compd. 1994; 216: 117-121.DOI: https://doi.org/10.1016/0925-8388(94)91052-9
4. Ren ZA, Lu W, Yang J, Yi W, Shen XL, Li ZC,Che GC, Dong XL, Sun LL, Zhou F, Zhao ZX. Superconductivity at 55 K in iron-based F-doped layered quaternary compound Sm [O1-x Fx] FeAs. Superconductivity Centennial 2019; 25: 223-227. DOI: https://doi.org/10.1142/9789813273146_0016
5. Zimmer BI, Jeitschko W, Albering JH, Glaum R, Reehuis M. The rate earth transition metal phosphide oxides LnFePO, LnRuPO and LnCoPO with ZrCuSiAs type structure. J. Alloys Compd. 1995; 229: 238-242. DOI: https://doi.org/10.1016/0925-8388(95)01672-4
6. Krellner C, Kini NS, Brüning EM, Koch K, Rosner H, Nicklas M,Baenitz M, Geibel, C. CeRuPO: A rare example of a ferromagnetic Kondo lattice. Phys. Rev. B. 2007; 76:104418. DOI: https://doi.org/10.1103/PhysRevB.76.104418
7. Nientiedt AT, Jeitschko W, Pollmeier PG, Brylak M. Quaternary equiatomic manganese pnictide oxides AMnPO (A= La-Nd, Sm, Gd-Dy), AMnAsO (A= Y, La-Nd, Sm, Gd-Dy, U), and AMnSbO (A= La-Nd, Sm, Gd) with ZrCuSiAs type structure. Z. Naturforsch. B, 1997; 52: 560-564. DOI: https://doi.org/10.1515/znb-1997-0504
8. Wollesen P, Kaiser JW, Jeitschko W. Quaternary Equiatomic

Compounds LnZnSbO (Ln= La-Nd, Sm) with ZrCuSiAs-Type Structure. Z. Naturforsch. B 1997; 52: 1467-1470. DOI: https://doi.org/10.1515/znb-1997-1205

9. Nientiedt AT, Jeitschko W. Equiatomic quaternary rare earth element zinc pnictide oxides RZnPO and RZnAsO. Inorg. Chem. 1998; 37: 386-389. DOI: https://doi.org/10.1021/ic971058q

10. Quebe P, Terbüchte LJ, Jeitschko W. Quaternary rare earth transition metal arsenide oxides RTAsO (T= Fe, Ru, Co) with ZrCuSiAs type structure. J. Alloys Compd. 2000; 302: 70-74. DOI: https://doi.org/10.1016/S0925-8388(99)00802-6

11. Palazzi M, Carcaly C, Flahaut J. Un nouveau conducteur ionique (LaO) AgS. J. Solid State Chem. 1980; 3:, 150-155. DOI: https://doi.org/10.1016/0022-4596(80)90487-9

12. Zhu WJ, Huang YZ, Wu F, Dong C, Chen H, Zhao ZX. Synthesis and crystal structure of barium copper fluochalcogenides:[BaCuFQ (Q= S, Se)]. Mater. Res. Bull.1994; 29: 505-508. DOI: https://doi.org/10.1016/0025-5408(94)90038-8

13. Kamihara Y, Hiramatsu H, Hirano M, Kawamura R, Yanagi H, Kamiya T, Hosono H. Iron-based layered superconductor: LaOFeP. J. Am. Chem. Soc. 2006; 128: 10012-10013. DOI: https://doi.org/10.1021/ja063355c

14. Welter R, Halich K, Malaman, B. Magnetic study of the ThCr2Si2-type RIr2Si2 (R= Pr, Nd) compounds: Magnetic structure of NdIr2Si2 from powder neutron diffraction. J. Alloys Compd. 2003; 353: 48-52. DOI: https://doi.org/10.1016/S0925-8388(02)01298-7

15. Kamihara Y, Hirano M, Yanagi H, Kamiya T, Saitoh Y, Ikenaga E, Kobayashi K, Hosono, H. Electromagnetic properties and electronic structure of the iron-based layered superconductor LaFePO. Phys. Rev. B, 2008; 77: 214515. DOI: https://doi.org/10.1103/PhysRevB.77.214515

16. Liang CY, Che RC, Yang HX, Tian HF, Xiao RJ, Lu JB, Li R,Li JQ. Synthesis and structural characterization of LaOFeP superconductors. Supercond. Sci. Technol. 2007, 20, 687. DOI: https://doi.org/10.1088/0953-2048/20/7/017

17. Kamihara Y, Watanabe T, Hirano M, Hosono, H. Iron-based layered

superconductor La [O1-x F x] FeAs (x= 0.05− 0.12) with T c= 26 K.J. Am. Chem. Soc. 2008; 130: 3296-3297. DOI: https://doi.org/10.1021/ja800073m

18. Katayama N, Kudo K, Onari S, Mizukami T, Sugawara K, Sugiyama Y, Kitahama Y, Iba K, Fujimura K, Nishimoto N, Nohara M. Superconductivity in Ca1-xLaxFeAs2: A Novel 112-Type Iron Pnictide with Arsenic Zigzag Bonds. J. Phys. Soc. Jpn. 2013;82(12):123702. DOI: https://doi.org/10.7566/JPSJ.82.123702

19. Yakita H, Ogino H, Okada T, Yamamoto A, Kishio K, Tohei T, Ikuhara Y, Gotoh Y, Fujihisa H, Kataoka K, Eisaki H. A new layered iron arsenide superconductor:(Ca, Pr) FeAs2. J. Am. Chem. Soc. 2014;136(3):846-9. DOI: https://doi.org/10.1021/ja410845b

20. Kudo K, Mizukami T, Kitahama Y, Mitsuoka D, Iba K, Fujimura K, Nishimoto N, Hiraoka Y, Nohara M. Enhanced Superconductivity up to 43 K by P/Sb Doping of Ca1− x La x FeAs2. J. Phys. Soc. Jpn. 2014;83(2):025001. DOI: https://doi.org/10.7566/JPSJ.83.025001

21. Kudo K, Kitahama Y, Fujimura K, Mizukami T, Ota H, Nohara M. Superconducting transition temperatures of up to 47 K from simultaneous rare-earth element and antimony doping of 112-type CaFeAs2. J. Phys. Soc. Jpn. 2014;83(9):093705. DOI: https://doi.org/10.7566/JPSJ.83.093705

22. Bert NA, Chaldyshev VV, Goloshchapov SI, Kozyrev SV, Kunitsyn AE, Tretyakov VV,Veinger AI, Ivonin IV, Lavrentieva LG, Vilisova MD, Yakubenya MP, Lubyshev DI, Preobrazhenskii VV, Semyagin B. Clusters and the Nature of Superconductivity in Ltmbe-GaAs. MRS Proceedings 1993; 325: 401. DOI: https://doi.org/10.1557/PROC-325-401

23. Chen XH, Wu T, Wu G, Liu RH, Chen H, Fang DF. Superconductivity at 43 K in SmFeAsO1-xF x. Nature, 2008; 453: 761-762. DOI: https://doi.org/10.1038/nature07045

24. Gordon RT. London penetration depth measurements in Barium (Iron1-xTx) 2Arsenic2 (T= Cobalt, Nickel, Ruthenium, Rhodium, Palladium, Platinum, Cobalt+ Copper) superconductors. Ph. D. Thesis 2011.

25. Kobayashi K, Yokoyama H. Superconductivity and antiferromagnetism in the phase diagram of the frustrated Hubbard model within a

variational study. Phys. C (Amsterdam, Neth.) 2010; 470: 1081-1084. DOI: https://doi.org/10.1016/j.physc.2010.05.041

26. Mounce AM, Oh S, Mukhopadhyay S, Halperin WP, Reyes AP, Kuhns PL, Fujita K, Ishikado M, Uchida S. Spin-Density Wave near the Vortex Cores in the High-Temperature Superconductor Bi2Sr2CaCu2O8+ y. Phys. Rev. Lett. 2011; 106: 057003. DOI: https://doi.org/10.1103/PhysRevLett.106.057003

27. Craco L, Laad MS. Normal state incoherent pseudogap in FeSe superconductor. Eur. Phys. J. B 2016; 89: 1-7. DOI: https://doi.org/10.1140/epjb/e2016-60842-y

28. Yamaji K, Yanagisawa T, Hase I. 3-Band theory of Fe pnictide superconductors. Phys. C (Amsterdam, Neth.) 2010; 470: 1060-1062. DOI: https://doi.org/10.1016/j.physc.2010.05.035

29. Suhl H, Matthias BT, Walker LR. Bardeen-Cooper-Schrieffer theory of superconductivity in the case of overlapping bands. Phys. Rev. Lett. 1959; 3: 552. DOI: https://doi.org/10.1103/PhysRevLett.3.552

30. Lesovik GB, Martin T, Blatter G. Electronic entanglement in the vicinity of a superconductor. Eur Phys J B 2001; 24: 287-290. DOI: https://doi.org/10.1007/s10051-001-8675-4

31. Johnston DC. Elaboration of the α-model derived from the BCS theory of superconductivity. Supercond Sci Technol. 2013; 26: 115011. DOI: https://doi.org/10.1088/0953-2048/26/11/115011

32. Li Z, Chen G, Dong J, Li G, Hu W, Wu D, Su S,Zheng P, Xiang T, Wang N, Luo J. Strong-coupling superconductivity in the nickel-based oxypnictide LaNiAsO 1− x F x. Phys. Rev. B: Condens. Matter 2008; 78: 060504. DOI: https://doi.org/10.1103/PhysRevB.78.060504

33. Illas F, Martin RL. Magnetic coupling in ionic solids studied by density functional theory. J. Chem. Phys. 1998; 108: 2519-2527. DOI: https://doi.org/10.1063/1.475636

34. Gyorffy BL, Staunton JB, Stocks GM. Fluctuations in density functional theory: random metallic alloys and itinerant paramagnets. In Density Functional Theory. Springer, Boston, MA, 1995;337: 461-484.

35. Kümmel S, Brack M. Quantum fluid dynamics from density-functional theory. Phys. Rev. A 2001; 64: 022506. DOI: https://doi.org/10.1103/PhysRevA.64.022506

36. Car R, Parrinello M. Unified approach for molecular dynamics and density-functional theory. Phys. Rev. Lett. 1985; 55: 2471. DOI: https://doi.org/10.1103/PhysRevLett.55.2471

37. Koskinen M, Lipas PO, Manninen M. Unrestricted shapes of light nuclei in the local-density approximation: comparison with jellium clusters. Nucl. Phys. A 1995; 591: 421-434. DOI: https://doi.org/10.1016/0375-9474(95)00209-J

38. Schmid RN, Engel E, Dreizler RM. Density functional approach to quantum hadrodynamics: Local exchange potential for nuclear structure calculations. Phys. Rev. C 1995; 52: 164. DOI: https://doi.org/10.1103/PhysRevC.52.164

39. Chen P, Qin M, Chen H, Yang C, Wang Y, Huang F. Cr incorporation in CuGaS2 chalcopyrite: A new intermediate-band photovoltaic material with wide-spectrum solar absorption. Phys. Status Solidi A 2013 210: 1098-1102. DOI: https://doi.org/10.1002/pssa.201228721

40. Palacios P, Sánchez K, Conesa JC, Wahnón P. First principles calculation of isolated intermediate bands formation in a transition metal-doped chalcopyrite-type semiconductor. Phys. Status Solidi A 2006; 203: 1395-1401. DOI: https://doi.org/10.1002/pssa.200566179

41. Palacios P, Sánchez K, Conesa JC, Fernández JJ, Wahnón P. Theoretical modelling of intermediate band solar cell materials based on metal-doped chalcopyrite compounds. Thin Solid Films 2007; 515: 6280-6284. DOI: https://doi.org/10.1016/j.tsf.2006.12.170

42. Ranjan P, Chakraborty T. Structure and optical properties of (CuAg) n (n= 1-6) nanoalloy clusters within density functional theory framework. J. Nanopart. Res. 2020; 22: 1-11. DOI: https://doi.org/10.1007/s11051-020-05016-0

43. Ranjan P, Dhail S, Venigalla S, Kumar A, Ledwani L, Chakraborty T. A theoretical analysis of bi-metallic (Cu-Ag)n=1-7 nano alloy clusters invoking DFT based descriptors. Mater. Sci.-Pol. 2015; 33: 719-724. DOI: https://doi.org/10.1515/msp-2015-0121

44. Ranjan P, Venigalla S, Kumar A, Chakraborty T. Theoretical Study Of Bi-Metallic Ag m Au n;(m+ n= 2-8) Nano Alloy Clusters In Terms Of Dft Based Descriptors, New Front. Chem, 2014; 23: 111-122.

45. Ranjan P, Chakraborty T. Density functional approach: to study copper

sulfide nanoalloy clusters. Acta Chim. Slov 2019; 66: 173-181. DOI: http://dx.doi.org/10.17344/acsi.2018.4762

46. Ranjan P, Chakraborty T, Kumar A. Density functional study of structures, stabilities and electronic properties of clusters: comparison with pure gold clusters. Mater. Sci.-Pol. 2020; 38: 97-107.DOI: https://doi.org/10.2478/msp-2020-0014

47. Das S, Ranjan P, Chakraborty T. Computational study of CunAgAu (n= 1-4) clusters invoking DFT based descriptors. Phys. Sci. Rev. 2022; DeGruyter 2022. DOI: https://doi.org/10.1515/psr-2021-0141

48. Ranjan P, Das S, Yadav P, Tandon H. Chaudhary S, Malik B, Rajak AK, Suhag V, Chakraborty T. Structure and electronic properties of [AunV] λ (n= 1-9; λ= 0,±1) nanoalloy clusters within density functional theory framework. Theor. Chem. Acc. 2021; 140: 1-12. DOI: https://doi.org/10.1007/s00214-021-02772-7

49. Das S, Chakraborty T, Ranjan P. Theoretical analysis of AgFen (n= 1-5) clusters: A DFT study. Mater. Today: Proc. 2022.54(3), 873-877

50. Gaussian 16, Revision C.01, Frisch MJ, Trucks GW, Schlegel HB, Scuseria GE, Robb MA, Cheeseman JR, Scalmani G, Barone V, Petersson GA, Nakatsuji H, Li X, Caricato M, Marenich AV, Bloino J, Janesko BG, Gomperts R, Mennucci B, Hratchian HP, Ortiz JV, Izmaylov AF, Sonnenberg JL, Williams-Young D, Ding F, Lipparini F, Egidi F, Goings J, Peng B, Petrone A, Henderson T, Ranasinghe D, Zakrzewski VG, Gao J, Rega N, Zheng G, Liang W, Hada M, Ehara M, Toyota K, Fukuda R, Hasegawa J, Ishida M, Nakajima T, Honda Y, Kitao O, Nakai H, Vreven T, Throssell K, Montgomery JA Jr, Peralta JE, Ogliaro F, Bearpark MJ, Heyd JJ, Brothers EN, Kudin KN, Staroverov VN, Keith TA, Kobayashi R, Normand J, Raghavachari K, Rendell AP, Burant JC, Iyengar SS, Tomasi J, Cossi M, Millam JM, Klene M, Adamo C, Cammi R, Ochterski JW, Martin RL, Morokuma K, Farkas O, Foresman JB. Fox DJ. Gaussian, Inc., Wallingford CT, 2016.

51. Parr RG, Yang W. Density-functional theory of atoms and molecules Oxford Univ. Press. ed: Oxford, 1989.

52. Shen M, Schaefer III HF. Dodecahedral and smaller arsenic clusters: As n, n= 2, 4, 12, 20. J. Chem. Phys. 1994; 101: 2261-2266. DOI: https://doi.org/10.1063/1.467666

53. Fujimoto H, Kato S, Yamabe S, Fukui, K. Molecular orbital calculations of the electronic structure of borazane. In Frontier Orbitals and Reaction Paths: Selected Papers of Kenichi Fukui, World Scientifuc Series in 20th Century Chemistry1997; 283-289.

54. Kato S, Fujimoto H, Yamabe S, Fukui, K. Molecular orbital calculation of the electronic structure of borane carbonyl. J. Am. Chem. Soc. 1974; 96: 2024-2029.DOI: https://doi.org/10.1021/ja00814a008

55. Ghosh DC, Bhattacharyya S. Molecular orbital and density functional study of the formation, charge transfer, bonding and the conformational isomerism of the boron trifluoride (BF3) and ammonia (NH3) donor-acceptor complex. Int. J. Mol. Sci. 2004; 5: 239-264. DOI: https://doi.org/10.3390/i5050239

56. Pati R, Senapati L, Ajayan PM, Nayak SK. First-principles calculations of spin-polarized electron transport in a molecular wire: Molecular spin valve. Phys. Rev. B 2003; 68: 100407-1-4. DOI: https://doi.org/10.1103/PhysRevB.68.100407

57. Ruiz-Morales Y. HOMO−LUMO Gap as an Index of Molecular Size and Structure for Polycyclic Aromatic Hydrocarbons (PAHs) and Asphaltenes: A Theoretical Study. I. J. Phys. Chem. A, 2002; 106: 11283-11308. DOI: https://doi.org/10.1021/jp021152e

58. Xiao H, Tahir-Kheli J, Goddard III WA. Accurate band gaps for semiconductors from density functional theory. J. Phys. Chem. Lett. 2011; 2: 212-217. DOI: https://doi.org/10.1021/jz101565j

59. Saravanan S, Balachandran V. Quantum chemical studies, natural bond orbital analysis and thermodynamic function of 2, 5-dichlorophenylisocyanate. Spectrochim. Acta, Part A 2014; 120: 351-364. DOI: https://doi.org/10.1016/j.saa.2013.10.042

60. Azam F,Alabdullah NH, Ehmedat HM, Abulifa AR, Taban I, Upadhyayula S. NSAIDs as potential treatment option for preventing amyloid β toxicity in Alzheimer's disease: an investigation by docking, molecular dynamics, and DFT studies. J. Biomol. Struct. Dyn. 2018; 36:2099-2117. DOI: https://doi.org/10.1080/07391102.2017.1338164

61. Parr RG, Zhou Z. Absolute hardness: unifying concept for identifying shells and subshells in nuclei, atoms, molecules, and metallic clusters. Acc. Chem. Res. 1993; 26: 256-258.DOI: https://doi.org/10.1021/ar00029a005

62. Chattaraj PK, Sengupta S. Chemical hardness as a possible diagnostic of the chaotic dynamics of rydberg atoms in an external field. J. Phys. Chem. A 1999; 103: 6122-6126. DOI: https://doi.org/10.1021/jp990242p

63. Pearson RG. Recent advances in the concept of hard and soft acids and bases. J. Chem. Educ. 1987; 64: 561. DOI: https://doi.org/10.1021/ed064p561

64. Sanderson RT. An interpretation of bond lengths and a classification of bonds. Science 1951; 114: 670-672. DOI: https://doi.org/10.1126/science.114.2973.670

65. Sanderson RT. Carbon—carbon bond lengths. Science 1952; 116: 41-42. DOI: https://doi.org/10.1126/science.116.3002.41

66. Parr RG, Szentpály LV, Liu S. Electrophilicity index. J. Am. Chem. Soc. 1999; 121: 1922-1924. DOI: https://doi.org/10.1021/ja983494x

Chapter - 6

Automated Detection and Classification of Mass Calcifications in Breast Cancer Mammograms using Image Processing Techniques

Author

Shreya Adhikary
Department of Electronics and Communication Engineering, Swami Vivekananda University, Barrackpore, West Bengal, India

Chapter - 6

Automated Detection and Classification of Mass Calcifications in Breast Cancer Mammograms using Image Processing Techniques

Shreya Adhikary

Abstract

Aspects of microscope imaging have already been investigated for the automatic identification of cervical cancer, while brightfield techniques have emerged for the detection of cervical cancer in fresh tissue. Using epithelial cells taken from the endo-cervical region, it might be a successful strategy. This study avoids the issue of nucleus segmentation and instead presents a technique for the automatic classification of cervical cells. This technique used a collection of spectral texture features (SPTF) that have been taken from single cell images that are transformed into two dimensions (2-D). Two 1-D functions, namely the radial function and the angular function accumulated from the frequency spectrum, are used to evaluate SPTFs in order to distinguish between normal and pathological cells. Support vector machines (SVMs) are used to evaluate the classification. The Gaussian SVM reaches to the greatest accuracy is 0.94, with sensitivity 0.89 and specificity 0.83. A promising performance more than 0.75 is obtained to achieve the total accuracy.

Keywords: Cervical cancer, human papillomavirus, spectral texture features, linear support vector machine, non-linear support vector machine.

Introduction

In 2023, cervical cancer ranked fourth globally among malignant tumors in women, accounting for 6.6% of all cancer diagnoses in females [1]. The cancer is developed by a sequence of alterations from precancer. However, it is generally found that, in 90% case, the early detection and treatment of precancerous lesions may prevent the individual cervical cancer. Human papillomavirus is responsible for this type of cancer.

Most of the proposed techniques employed the methods that are subject to do segmentation of the cytoplasm and nucleus [2-3]. To circumvent the problem of cell and nuclei segmentation, several attempts have focused on the categorization of the Papanicolaou (Pap) test using sets of properties derived from the frequency domain. The virtue of this approach is that rigorous segmentation of the cell and its nucleus is not elementary. However, it should be noted that this approach was only applied to the single-cell image captured at high resolution.

This work recommends a new procedure to classify traditional cervical cells and abnormal cells (precancerous cells) by utilizing a set of basic properties extracted from Fourier transform of low- resolution cell images. This allows us to bypass the cell segmentation process and a high-resolution image purchasing system.

Methodology

This section represents the method of this research work. The performance of this work is evaluated by using a publicly available dataset of 800 cervical Pap smear cell images, collected from HErlev Pap Smear Dataset [4]. The dataset are visualized in Fig. 1.

Firstly, the DFT transforms M×N gray scale image into 2-D Fourier transform, transforming components into polar coordinates and separated into radial and angular functions. All elements of the Fourier spectrum are transformed into polar coordinates, F(r, θ), which are divided into two 1-D functionalities, that is, the radial functionality, F(r) and the angular functionality, F(θ) [5].

Radial functionality and angular functionality

The spectrum consists of frequency components along a circle of radius r and a radial line with angle θ, represented by mean, variance, and entropy through statistical evaluations. The features of normal and abnormal cells have been calculated using annular ring subsections and radian wedge subsections. The image information is divided into a frequency component with a small annular ring, divided into 8 equidistance annular rings with mean, variance, and entropy. The cervical dataset is utilized by two labels, i.e., normal and abnormal. Normal data represents the healthy volunteer and abnormal data specifies diseased person with cancer. This work presents machine learning approaches to predict and classify cervical cells. It includes stages of data collection, model training, testing, and comparing outcomes.

The dataset is divided into two parts as training and testing part with 80% and 20% respectively.

Vapnic invented the Support Vector Machine (SVM) to solve data classification and regression problems by dividing original data into different groups and creating a hyperplane to predict new data labeling [6]. The SVM algorithm aims to construct the optimum decision boundary or line that can divide n-dimensional space into classes so that we can quickly classify the data points. The optimal decision boundary, called hyperplane, which is separated the two types of data. The hyperplane is created by SVM by selecting the extreme points and vectors. The SVM classifier is based on these extreme situations, which are referred to as support vectors. The SVM algorithm is often reported to achieve better results than other classifiers, although it has been indicated that the main reason to use an SVM instead is because the problem might not be linearly separable. In that case, an SVM with a non-linear kernel such as the Radial Basis Function (RBF) would be suitable [7].

Performance measures

The classification is analyzed by some statistical methods are described below:

Sensitivity: The fraction of diseased cases predicted as diseased

$$Sensitivity = \frac{TP}{TP + FN}$$

Specificity: The fraction of normal cases predicted as normal one

$$Specificity = \frac{TN}{TN + FP}$$

Accuracy: The fraction of cases the model correctly identified

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

Where, True Positive (TP): Predicted diseased, actual diseased

False Positive (FP): Predicted diseased, actual non-diseased False Negative (FN): Predicted non-diseased, actual diseased True Negative (TN): Predicted non-diseased, actual non-diseased

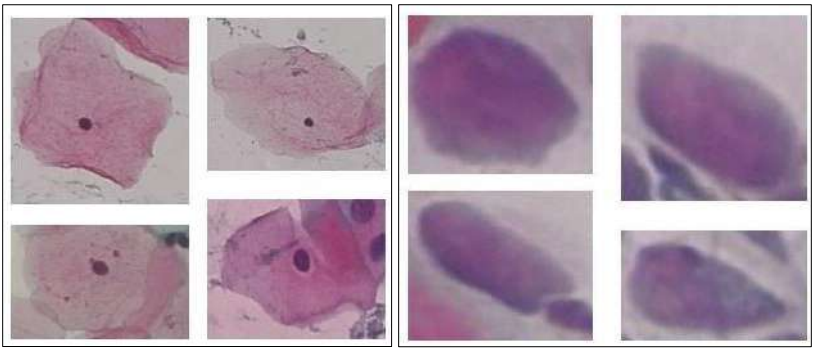


Figure 1: HErlev dataset: normal (left column) and cancer (right column)

Results and discussions

This section presents two SVM-based approaches for classification of cervical cancer data, evaluated using 10-fold cross-validation. Two target variables are diagnosed, and the total accuracy of these methods is plotted.

The non-linear SVM can classify the abnormal dataset and achieve high classification accuracy of 0.94. The linear SVM and non-linear SVM reduce computation costs and can perform diagnosis with less factors or components. Though, the non-linear SVM method effectively detects faulty samples, with a total accuracy of 0.94 for cervical cancer dataset. Classification results of HErlev Pap test dataset using linear and non-linear SVM techniques are graphically shown in fig. 2 and also depicted in tabular form in Table 1.

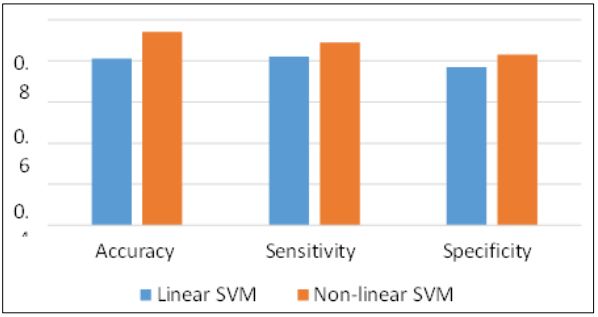


Figure 2: The classification results of the present study with respect to accuracy, sensitivity and specificity

Table 1: Classification performance of HErlev Pap test dataset using linear and non-linear dataset

Classifiers	Accuracy	Sensitivity	Specificity
Linear SVM	0.81	0.82	0.77
Non-linear SVM	0.94	0.89	0.83

Conclusion

This paper shows the cervical cancer risk factors and compares three SVM-based approaches for classification of cervical cancer datasets. Linear SVM and Non-linear SVM effectively classify malignant and benign cancers. With linear SVM shows improved classification results. However, linear SVM has better capability with the same number of features. This research study investigates the classification of normal and abnormal cervical cells using linear and non-linear SVM classifiers with simplified set of SPTFs. Linear SVM has low classification performance, while non-linear SVM achieves highest accuracy and specificity.

References

1. https://www.who.int/health-topics/cervical-cancer#tab=tab_1
2. J. V. Lorenzo-Ginori, W. Curbelo-Jardines, J. D. L´opez-Cabrera, and S.B. Huergo-Su´arez, Cervical cell classification using features related to morphometry and texture of nuclei, Lecture Notes in Computer Science, vol. 8259 LNCS, no. PART 2, pp. 222-229 (2013)
3. M. E. Plissiti, C. Nikou, and A. Charchanti, Combining shape, texture and intensity features for cell nuclei extraction in Pap smear images, Pattern Recognition Letters, vol. 32, no. 6, pp. 838-853 (2011)
4. https://paperswithcode.com/dataset/herlev
5. T. Chankong, Automatic Classifying of Cervical Cells Using Fourier Spectral Features, Proc. 2018 4th Int. Conf. Green Technol. Sustain. Dev. GTSD, pp. 759-762 (2018)
6. V. Cortes, Corinna and Vapnik, Photonit neural networks and learning machines the role of electron-trapping materials, IEEE Expert, vol. 7, no. 5, pp. 63-72(1992)
7. https://pubmed.ncbi.nlm.nih.gov/9691570

Chapter - 7

The Emergence of Periodic and Chaotic Patterns in a Simple Predator -Prey System in Presence of Self-Diffusion

Authors

Santu Ghorai
Department of Basic Science and Humanities, University of Engineering and Management, India

Shubham Shaw
Deloitte, Senapati Bapat Marg, Fitwala Road, Mumbai, Maharashtra, India

Subabrata Mondal
Department of Mathematics, Swami Vivekananda University, India

Subhendu Maji
Department of Mathematics, Swami Vivekananda University, India

Chapter - 7

The Emergence of Periodic and Chaotic Patterns in a Simple Predator -Prey System in Presence of Self-Diffusion

Santu Ghorai, Shubham Shaw, Subabrata Mondal and and Subhendu Maji

Abstract

The research investigates the temporal and spatial dynamics of a basic predator-prey system featuring a Holling Type II functional response and incorporating self-diffusion. Initially, we considered a general linear system and derived conditions that determine stability in both Hopf and Turing patterns within a spatial context through linear stability analysis. It was observed that the Turing pattern does not emerge in this system. To validate our analytical results, we conducted numerical simulations. These simulations were performed with zero flux boundary conditions around a steady state. They revealed various spatial patterns, including oscillations and spatiotemporal chaos when the predator diffusion is increased. These investigations offer valuable insights into the complex spatiotemporal dynamics of population models.

Keywords: Diffusion, dispersion, instability, periodic and chaotic pattern formation.

Introduction

In our daily lives, mathematics plays a remarkable and often hidden role in understanding the myriad of phenomena we encounter. Whether these phenomena are readily connected to mathematics or seem unrelated at first glance, we can discern the pervasive influence of mathematical principles upon closer examination. This ubiquity of mathematics extends its reach across both micro and macro scales, providing a powerful tool for articulating and comprehending these phenomena through mathematical models - a practice commonly referred to as mathematical modeling. Remarkably, these mathematical models have the capability to encompass a wide range of entities, from celestial bodies like planets and stars to the microscopic world of viruses and bacteria.

Within scientific inquiry, ecology is a discipline that explores the intricate interrelationships among organisms and their environment. This multifaceted field encompasses various subjects, including the description of species interactions, the conservation of species, the development of ecosystems, the study of biodiversity, and the analysis of disease spread across different species. Broadly speaking, the study of ecology is divided into two principal branches: experimental ecology, which investigates ecological systems through controlled experiments, and mathematical ecology, which harnesses mathematical models to address complex ecological questions related to biodiversity, epidemiology, and other ecological concerns. The results of mathematical ecology are then translated into ecological terms, contributing significantly to our understanding of ecological processes.

One of the enduring and well-studied themes in ecology revolves around predator-prey dynamics, a concept that has garnered significant attention since the pioneering works of Alfred Lotka and Vito Volterra [1, 2]. These dynamics delve into crucial components, including the growth function of prey species, the mortality function of predator species, and the nature of their interactions, often captured through functional responses. Mathematical ecologists employ various growth functions, such as Malthusian and logistic growth, and different functional responses, including Holling types I-IV and the Beddington-DeAngelis functional response. The choice of these functions depends on the specific species and ecological interactions under consideration. Our natural surroundings surround us with an abundance of patterns, some readily observable with the naked eye, while others remain concealed unless closely examined. These patterns have captivated human curiosity for centuries. The diversity of profound patterns in everyday life, including the intricate morphological shapes of living organisms (both plants and animals), the formation of clouds, the symmetry of snowflakes, and the winding paths of rivers, cannot be overstated. These spatiotemporal patterns are ubiquitous, leading to the emergence of a relatively new research field known as pattern formation. Alan Turing, a renowned mathematician and pioneer in computer science, significantly contributed to the field of pattern formation with his seminal 1952 paper titled "The Chemical Basis of Morphogenesis" [3]. Turing's groundbreaking work elucidated the mechanisms underlying pattern formation and emphasized the role of diffusion, the net movement of molecules or particles driven by concentration gradients. According to Turing, diffusion can disrupt the stable

equilibrium of a system, leading to what he termed "diffusion-driven instability." His work remains highly relevant today, as Turing instability is instrumental in generating spatially organized patterns in a wide range of disciplines. Turing posited that a reaction-diffusion system should involve at least two reactive species diffusing at significantly different rates to generate spatial patterns. This concept has applications in diverse fields, including chemical reactions, astrophysics, hydrodynamics, ecological systems, and economics. The mechanisms Turing proposed for pattern formation continue to be explored and refined in contemporary research.

While diffusion is commonly associated with the movement of molecules or small particles, it is important to note that it can also occur at macroscopic levels within ecological systems. Species' movement to more suitable habitats, essential for their survival and presence, is often driven by various factors, including predator-prey relationships, overcrowding, climate variations, refuge-seeking behaviors, and resource limitations within a given habitat. In this context, diffusion can be conceptualized as the propensity of a biological population initially concentrated at a point in ecological space to spread over a larger area over time. This process increases system stability by reducing local overcrowding and resource competition. Pioneers in the field, including Philip Maini, Michael Murray, and John Tyson, introduced reaction-diffusion concepts to ecology and applied Turing's ideas of pattern formation to predator-prey population dynamics [4,5]. They argued that diffusion should be perceived as the dispersal of population density, and it is often considered a stabilizing process that leads to the formation of spatial patterns. Specifically, when diffusion involves the movement of a particular species due to a concentration gradient of the same species, it is referred to as self-diffusion.

In the ensuing sections of this article, we embark on a journey to establish a general diffusion model and subsequently apply it to a specific predator-prey system. Our analysis will delve into the effects of varying the diffusion rate of predators. The predator-prey system under consideration exhibits logistic growth of prey, a Holling type II functional response, and a linear death rate of predators.

General dynamics in the presence of self-diffusion

Let us consider a general two-species predator-prey system. Let, at time t are $n(t)$ and $p(t)$ are the biomass of prey and predator, respectively

$$\frac{dn}{dt} = f(n,p), \frac{dp}{dt} = g(n,p) \quad (1)$$

In the presence of self-diffusion of the model (1) can be written as

$$\frac{\partial n}{\partial t} = f(n,p) + D_1 \nabla^2 n \ \frac{\partial p}{\partial t} = g(n,p) + D_2 \nabla^2 p, \quad (2)$$

where the coefficients $D_1, and\ D_1$ represent self-diffusion coefficients for prey and predator species, respectively and are always positive because they move from higher to lower concentration. $\nabla^2 =$

$\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}$ is the Laplacian operator. We wish to investigate model (2) under the positive initial conditions, i.e.,

$$n(\vec{x}, 0) > 0, p(\vec{x}, 0) > 0, \vec{x} = (x, y) \in \Omega \subset R^2 \quad (3)$$

with no-flux boundary conditions

$$\frac{\partial n}{\partial \nu} = \frac{\partial p}{\partial \nu} = 0, on\ \partial\Omega \times (0, t] \quad (4)$$

where Ω is the two-dimensional spatial domain, and ν is the outward unit normal vector at the boundary $\partial\Omega$. The equation (4) implies that species in Ω cannot leave outside of the domain Ω, and the species outside cannot enter the domain Ω.

Bifurcation analysis

Local dynamics

Without loss of generality, let $E^* \equiv (n^*, p^*)$ be the positive equilibrium point. The characteristic equation of the system without diffusion i.e., system (1), is given by $\lambda^2 - tr(J_0)\lambda + \Delta_0 = 0$ where $tr(J_0) = a_{11} + a_{22}, \Delta_0 = a_{11}a_{22} - a_{12}a_{21}$. For the local system to be stable, we must have

$$tr(J_0) < 0\ and\ \Delta_0 > 0, i.e., a_{11} + a_{22} < 0\ and\ a_{11}a_{22} - a_{12}a_{21} > 0. \quad (5)$$

2.1.2 Diffusion system analysis

Now, we will linearize system (2) around E^*. To do this, we will take the spatiotemporal perturbation around the equilibrium point E^* as

$$n(\vec{x}, t) = n^* + c_1 e^{\lambda t} e^{i\vec{k}\vec{x}}, p(\vec{x}, t) = p^* + c_2 e^{\lambda t} e^{i\vec{k}\vec{x}},$$

where $c_1 e^{\lambda t} e^{i\vec{k}\vec{x}} = \Delta n <<< 1$ and $c_2 e^{\lambda t} e^{i\vec{k}\vec{x}} = \Delta p <<< 1. c_1, c_2$ are constants, λ is the frequency, $\vec{k}$ is the wave number vector and $\vec{x}$ is the position vector. Following, Ghorai and Poria [6], the characteristic equation in the presence of spatiotemporal perturbation, reads $\lambda^2 - tr(J_k)\lambda + \Delta_k = 0$,

$$where,\ tr(J_k) = tr(J_0) - k^2(D_1 + D_2), \Delta_k = D_1 D_2 k^4 - (D_2 a_{11} + D_1 a_{22})k^2 + \Delta_0.$$

Turing bifurcation

For the Turing pattern, the local system must be stable, and there must be diffusion-driven instability [6]. To make system (2) unstable, we must have $\Delta_k < 0$.

The conditions for the Turing pattern can be summarized as follows:

$$(i)\ a_{11} + a_{22} < 0\ (6a)\ (ii)\ a_{11}a_{22} - a_{12}a_{21} > 0\ (6b)\ (iii)\ D_2 a_{11} + D_1 a_{22} > 0\ (6c)\ (iv)\ D_2 a_{11} + D_1 a_{22} > 2\sqrt{D_1 D_2 \Delta_0}\ (6d)$$

Application

Now, we are taking a particular form of $f(n,p)$ and $g(n,p)$ and analyse it under self-diffusion in similar way as done in the previous section and checking whether we obtain Turing pattern or not. We will consider a predator-prey system having Holling type-II functional response

$$\frac{dn}{dt} = rn\left(1 - \frac{n}{k}\right) - \frac{anp}{b+n}, \frac{dp}{dt} = \frac{canp}{b+n} - dp \quad (7)$$

where $n(t)$ and $p(t)$ are the prey and predator densities at time t, respectively. r is the specific growth rate of prey, k is the carrying capacity, a is the predation coefficient, b is the half-saturation coefficient, c is the conversion rate of prey into the predator, and d is the death rate of predator in absence of prey.

For convenience, we will transform system (7) into a non-dimensional form by doing the following transformations.

$$\delta = \frac{1}{r}\alpha = k, \beta = \frac{b}{\delta a} = \frac{br}{a} N = \frac{n}{k}, P = \frac{pa}{br}, T = tr$$

Putting the above relation in equation (7), we get

$$\frac{dN}{dT} = N(1-N) - \frac{NP}{1+BN}, (8)\ \frac{dP}{dT} = \frac{CNP}{1+BN} - DP$$

where, $B = \frac{k}{b}, D = \frac{d}{r}, C = \frac{cak}{br}$.

3.1. Bifurcation analysis

The interior equilibrium point $E^* \equiv (N^*, P^*)$, where

$$N^* = \frac{D}{C-BD}\ and\ P^* = \frac{C(C-BD-D)}{(C-BD)^2}. \quad (9)$$

We therefore, the Jacobian component the is

$$a_{11} = \frac{D[BC - C - B^2D - BD]}{C(C - BD)}, a_{12} = -\frac{D}{C}, a_{21} = C - BD - D, a_{22} = 0.$$

Thus,

$$tr(J_0) = \frac{D[C(B-1)-BD(B+1)]}{C(C-BD)}, \Delta_0 = \frac{D}{C}(C - BD - D). \quad (10)$$

It is observed that due to the absence of $a_{22}(= 0)$, the condition $a_{11} < 0$ and $D_1 a_{11} > 0$ cannot hold simultaneously. Thus, the Turing pattern cannot occur for this kind of system. However, it is noted that the Hopf and Hopf Turing pattern may exist. On the other hand, it is observed that

If $B \leq 1$, the local system (8) is globally asymptotically stable and if $B > 1$, the local system (8) will be stable if $D > D_H$, where D_H is the Hopf-bifurcation parameter given by the following expression

$$tr(J_0) < 0 \Rightarrow D_H < D, \left[where, D_H = \frac{C(B-1)}{B(B+1)} \right] \quad (11)$$

Therefore, Turing pattern is impossible in this predator-prey system with self-diffusion.

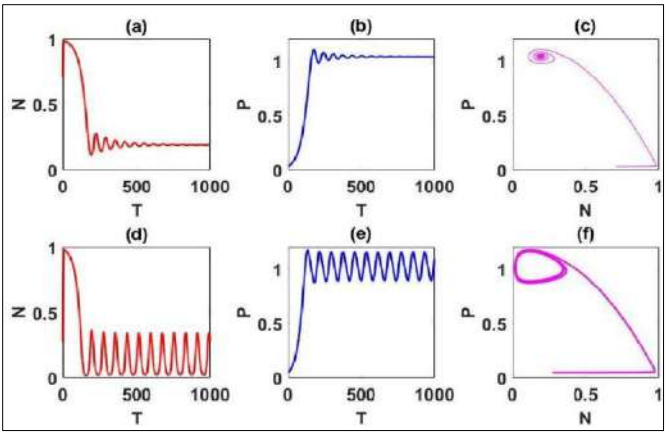


Figure 1: (a, d): The time series behavior of N of non-diffusive system; (b, e): The time series behavior of P of non-diffusive system. (c, f): Represent the corresponding phase portraits. The first row represents the stable behavior for $D = 0.015$, whereas the second row represents the unstable behavior of the non-diffusive system for $D = 0.01$. The other parameters are fixed as $B = 1.5$ and $C = 0.1$

Numerical results

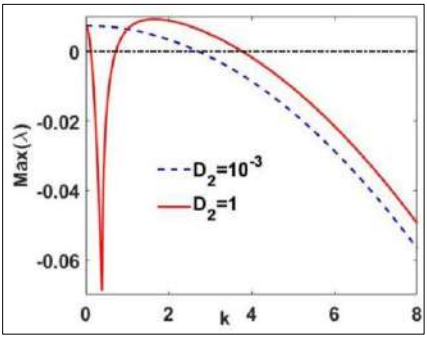


Figure 2: The dispersion relation of the system is depicted in the graph, where the blue dashed line represents the behavior associated with Hopf bifurcation for $D_2 = 10^{-3}$, while the solid red curve represents the behavior associated with Hopf-Turing for $D_2 = 1$. The other parameters are held fixed at $B = 1.5, C = 0.1, D = 0.01$, and $D_1 = 10^{-3}$.

Following Ghorai and Poria [6], we transform the continuous model to a finite dimensional model i.e., discretized the model in time and space.

We used the FTCS scheme for discretization and random initial conditions around the steady state. We take $L = 50$ the time step $\Delta t = 0.01$ and the step length $\Delta x = \Delta y = 0.5$. We fixed the other local parameters as $B = 1.5, C = 0.1, D = 0.01, D_1 = 10^{-3}$and taking D_2 as a control parameter.

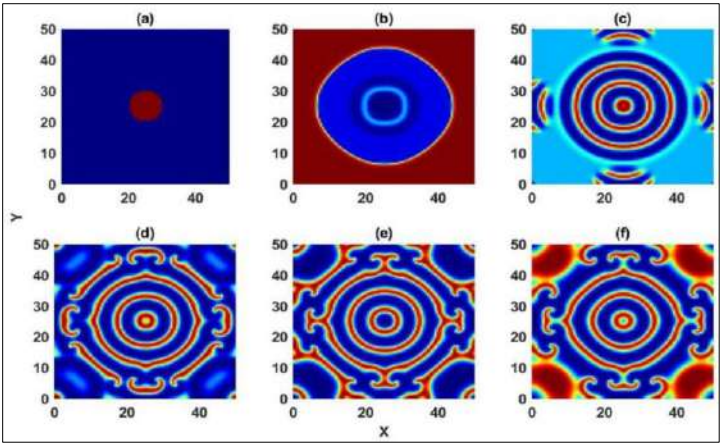


Figure 3: The evaluation of patterns at different time instants is conducted for $D_2 = 10^{-3}$, with the other parameters held fixed at $B = 1.5, C = 0.1, D = 0.01$, and $D_1 = 10^{-3}$

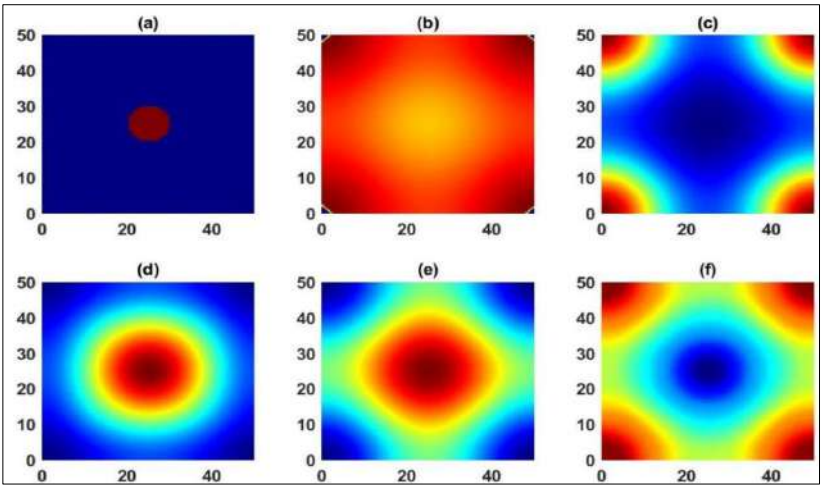


Figure 4: The evaluation of patterns at different time instants is conducted for $D_2 = 1$, with the other parameters held fixed at $B = 1.5, C = 0.1, D = 0.01$, and $D_1 = 10^{-3}$

In Fig. 1 we have shown the stable (upper row]) and unstable behavior (lower row) of the of the system without diffusion for $D = 0.015$ and $D = 0.01$. In Fig. 2, We have plotted the dispersion relation for two different values of D_2. The blue dash line represented the behavior of Hopf bifurcation for $D_2 = 10^{-3}$ and the solid red curve is represent the behavior of Hopf-Turing for $D_2 = 1$ The In 3; we have depicted the pattern evaluation of N species for a different instant of time considering the form the Hopf domain ($B = 1.5, C = 0.1, D = 0.01, D_1 = 10^{-3}$ and $D_2 = 10^{-3}$). A similar presentation is also shown in Fig. 4, when we choose the parameter from Hopf-Turing domain ($B = 1.5, C = 0.1, D = 0.01, D_1 = 10^{-3}$ and $D_2 = 1$). However, in the Hopf region, we have observed the periodic pattern, whereas in the Hopf-Turing area, we have observed the chaotic way, which are shown in Fig. 5.

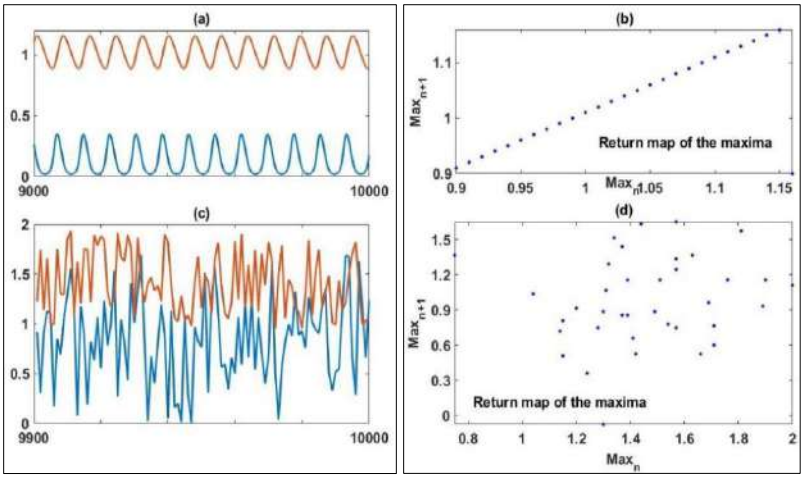


Figure 5: (a, c): The time series behavior of N and P for $D_2 = 10^{-3}$ and $D_2 = 1$, respectively; (b, d): Represent the Poincaré maximum return map. In (a, b), periodic behavior is shown, while chaotic behavior is depicted in (c, d). The other parameters are held fixed at $B = 1.5$ and $C = 0.1$

Summary

In this article, we study a relatively simple predator-prey system following a Holling-type II functional response model with linear predator death. What makes this research unique is that we introduce an element of self-propagation into this ecological network. Our first approach is a theoretical analysis of the system, where we investigate the basic notions of Turing and Hopf instability.

For Turing, instability is the phenomenon in which spatial patterns such as lines or lines appear in systems due to instabilities due to diffusion; in other words, Hopf instability is the origin of oscillatory behavior in a connected system. Our analysis revealed an interesting result: the absence of a particular specified term a_{22} in our model prevents the construction of Turing models. Consequently, we check that modeled models are only accessible to our parameter's Hopf and Hopf-Turing regions. We performed statistical simulations to confirm the results of our theoretical analysis. These simulations used favorable initial conditions and zero-flux boundary conditions to simulate real-world scenarios accurately. Our numerical experiments provided valuable insights. In areas of Hopf instability, we observed the emergence of periodic structures in the predator system. In contrast, in the Hopf-Turing region, we observed the formation of chaotic

systems by varying the dispersal rate of predators. In future research, it would be helpful to examine how changes in other parameters of the model affect the formation of this model. This may provide a deeper understanding of the above dynamics.

References

1. Lotka, A. J.: Elements of physical biology.
2. Volterra, V.: Variazione e fluttuazioni del numero d'individui in specie animali conviventi. Mem R Accad Naz dei Lincei, 2, 31-113 (1926).
3. Turing, A. M.: On the chemical basis of morphogenesis. Philos Trans Roy Soc London B, 237, 37-72 (1952).
4. Maini, P. K., Murray, J. D., & Othmer, H. G.: The spatial spread of animal populations: Patterns, problems, and implications. Journal of Theoretical Biology, 145(4), 511-525 (1989).
5. Tyson, J. J., & Murray, J. D.: Cyclic AMP waves during aggregation of Dictyostelium amoebae. Development, 106(3), 421-426 (1989).
6. Ghorai, S., & Poria, S.: Turing patterns induced by cross-diffusion in a predator-prey system in presence of habitat complexity. Chaos, Solitons & Fractals, 91, 421-429 (2016).

Chapter - 8

Dispersion Equation & it's Solution for Single Layer Fluid with Surface Tension Having an Inertial Surface

Author

Arijit Das
Department of Applied Mathematics, Swami Vivekananda University, Kolkata, West Bengal, India

Chapter - 8

Dispersion Equation & it's Solution for Single Layer Fluid with Surface Tension Having an Inertial Surface

Arijit Das

Abstract

This manuscript aims to determine the characteristics of the solutions to a dispersion equation that emerges during the examination of small-amplitude surface waves when a fluid with surface tension is covered by an inertial surface. In this scenario, the fluid possesses an inertial surface at the top, while at the bottom it has a rigid surface. By analyzing graphs of relevant functions, all solutions to this dispersion equation are identified and analyzed.

Keywords: Dispersion equation, surface tension, inertial surface, graphical method.

Introduction

For several decades, researchers have typically examined different categories of low-amplitude surface waves in water using the linearized theory of water waves (c.f. Wehausen and Laitone (1960)). Since gravity is the sole external force involved, these waves are commonly referred to as surface gravity waves. The relationship between the wave number and the angular frequency of such waves is termed the dispersion equation. This terminology arises from the phenomenon that waves with varying wavelengths, propagating at distinct velocities, tend to disperse. This dispersion occurs because the speed of wave propagation equals the ratio of the angular frequency to the wave number. In the case of infinitely deep water, if "k" represents the wave number and "ω" stands for the angular frequency of a surface gravity wave, the dispersion equation can be expressed as follows:

$$k = \frac{\omega^2}{g} \ (\equiv K) \qquad (1.1)$$

where g is the acceleration due to gravity. The solution of Equation (1.1) is k = K, and this corresponds to the time-harmonic progressive surface waves represented by $(x, y) = e^{\{-kyikx\}}$ where $Re\{(x, y)e^{\{-i\omega t\}}\}$ is the velocity potential describing the two-dimensional motion in deep water occupying the position $y \geq 0$ where the y-axis directed vertically downward, the plane z = 0 defines the mean free surface, and the x-direction represents the direction of wave propagation. For water of uniform finite depth h, Eq. (1.1) modifies to the following equation:

$$k \tan(kh) = K. \tag{1.2}$$

The equation (1.2) is the relation between the wave number and the angular frequency of train of surface gravity waves and is known as the dispersion equation for water of uniform finite depth. The term is due to the fact that waves whose velocity depends on wave number disperse or separate. Wave dispersion is a fundamental process in many physical phenomena. This dispersion relation $k \tanh(kh) = K$ has roots $\pm k_0$ and $\pm k_n$ (n = 1,2,3,--------, $k_1 < k_2 < k_3 <$----------) and there is no other root.

It is interesting to observe that small time-harmonic progressive gravity waves with a given angular frequency cannot exist at the surface of an ideal liquid, such as water, when it's covered by a thin uniform layer of floating matter (such as broken ice or unstretched mat) if this layer, referred to as the inertial surface, is too heavy. This outcome sharply contrasts with the situation where the inertial surface is sufficiently light, similar to the familiar free surface, allowing it to support such waves.

These two scenarios involving an inertial surface were noted and demonstrated in studies on waves occurring at a free surface adjacent to an inertial surface. Peters (1950) conducted research on this topic for cases of infinite depth, while Weitz and Keller (1950) explored it for situations with finite and constant depth. There is an ample amount of studies considering an inertial surface (see, Rhodes-Robinson (1982, 1983), Mandal and Kundu (1986)). In this present note we will study the nature of roots of the dispersion equation that arises when the free-surface is replaced by an inertial surface.

Formulation

Although the basic equations for a fluid with inertial surface are well known in the literature, but for the sake of completeness we briefly mention the equations here. The assumptions made in this context are as follows: It is

presumed that water behaves as an inviscid, incompressible, and uniform fluid. The motion within it is solely influenced by gravity and is characterized as irrotational. Furthermore, the assumption regarding the smallness of motion implies that the velocity components, along with their respective derivatives, are considered to be of a first-order small magnitude. Consequently, higher-order terms, including their squares, products, and other powers, can be disregarded in the analysis. A rectangular Cartesian coordinate system is chosen, with the y-axis directed vertically downward. The plane defined by $y = 0$ corresponds to the location of the mean free surface.

The equation of continuity is given by

$$^2\phi = 0, \text{ in the fluid region} \quad (2.1)$$

The condition of no motion at the bottom surface gives,

$$\overrightarrow{\phi} \to 0 \text{ as } y \to \infty, \text{ for deep water} \quad (2.2)$$

$$\frac{}{\partial y} = 0 \text{ on } y = h \text{ for water of uniform finite depth} \quad (2.3)$$

The condition at the inertial surface with surface tension yields (c.f. Rhodes-Robinson (1982))

$$K\,\varphi + (1 - \varepsilon\, K)\varphi_y + M\,\varphi_{yyy} = 0 \quad (2.4)$$

where $K = \frac{\omega^2}{g}$, $\varepsilon = \frac{m}{\rho\, g}$, and $M = \frac{T}{\rho\, g}$. Here, ρ is the density of water, m is the surface density of the inertial surface, T is the coefficient of surface tension, and ε is dimensional quantity. In is important to note here that, $\varepsilon = 0$ correspond to the water with free surface.

Solution of the dispersion equation

Now, we use the method of separation of variables to solve the boundary value problem. Let $\varphi(x,y) = X(x)\,Y(y)$

From(2.1), $\frac{^2}{\partial x^2} + \frac{^2}{\partial y^2} = 0$

or, $X''(x)\,Y(y) + X(x)\,Y''(y) = 0$

or, $\frac{X''(x)}{X(x)} = -\frac{Y''(y)}{Y(y)} = \text{constant } (= -k^2)$ (say)

The progressive wave solution of the problem for an infinite depth fluid is

$(x, y) = e^{\{-kyikx\}}$

Where the wave number k satisfies the following dispersion equation

$$M\,k^3 + k(1 - \varepsilon\,K) - K = 0. \qquad (3.1)$$

To analyze the nature of the roots of this dispersion equation let us choose

$$f(k) = M\,k^3 + k(1 - \varepsilon\,K) - K.$$

Clearly, f(∞)>0 but f(0)<0.

This shows that whether Kε > or < 1, there always exists one real positive root of (3.1) so long M>0. Thus, the existence of progressive wave is guaranteed.

For water of finite depth 'h' covered with I.S in presence of surface tension, the progressive wave solution is

$(x, y) = cosh\ cosh\ k(h - y)\ e^{\{ikx\}}$

Where the wave number k satisfies

$$K - k\{(1 - \varepsilon K) + M\,k^2\}\,tanh\ tanh\ kh\ = 0. \qquad (3.2)$$

To derive the nature of solution of the equation (2.4), we will use the graphical method.

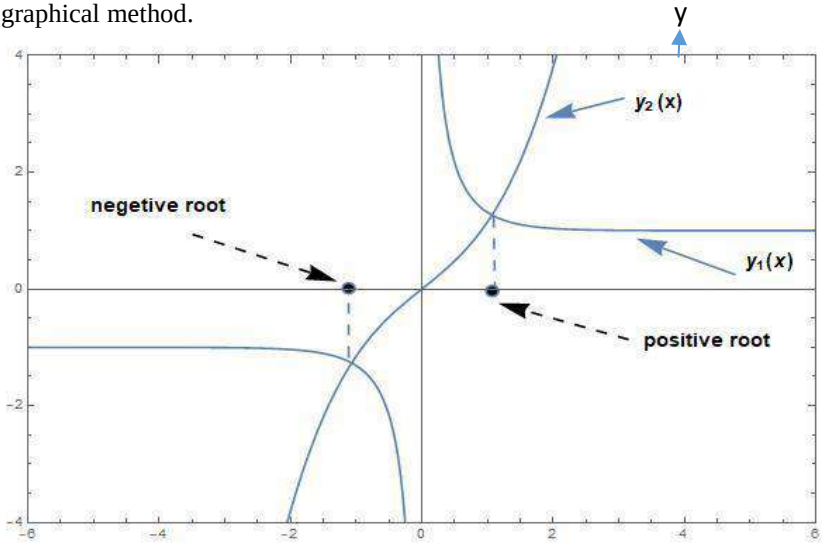


Figure 1: Nature of the real roots when εK<1

Graphical method and discussion

The equation (3.2) can be written as

$$coth\ coth\ x\ = \frac{x}{Kh}[(1-\varepsilon\ K)+\frac{M}{h^2}\ x^2] \tag{3.3}$$

where $x = k\ h$.

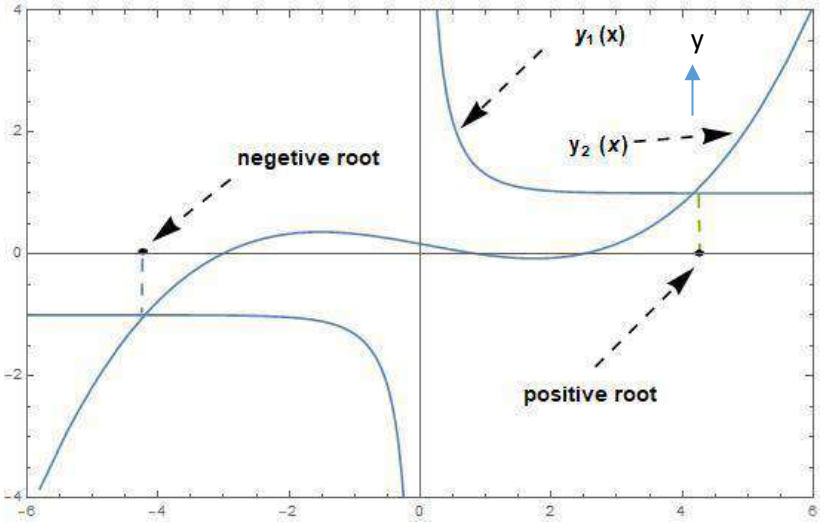


Figure 2: Nature of the real roots when εK>1

Let $y_1(x) = coth\ coth\ x$ and $y_2(x) = \frac{x}{Kh}[(1-\varepsilon\ K)+\frac{M}{h^2}\ x^2]$.

Then, $y_2(x) \to\ \infty\ as\ x \to\ +\infty$ and $y_2(x) \to\ -\infty\ as\ x \to\ -\infty$.

Clearly, the dispersion equation has at least one real root for any value of εK. The same is represented by figure 1 and figure 2.

It can be seen that though the nature of the curve $y_2(x)$ changes when εK>1, it still intersects $y_1(x) = coth\ coth\ x$ at only two points, one is positive and the other being negative ($\pm\alpha_0$).

To find the imaginary roots we change k by $i\ k$ and thus the equation changes to

$$cot\ cot\ kh = \frac{k}{K}[m\ k^2 - (1-\varepsilon K)]. \tag{3.3}$$

Putting $kh = x$, we get,

$$cot\ cot\ x = \frac{x}{kh}[\frac{M}{h^2}x^2 - (1-\varepsilon K)]. \tag{3.4}$$

As before we analyze the nature of the roots by solving equation (3.4) graphically.

Taking, $y_1(x) = cot\ cot\ x$ and $y_2(x) = \frac{x}{Kh}[\frac{M}{h^2}\ x^2 - (1 - \varepsilon\ K)]$.

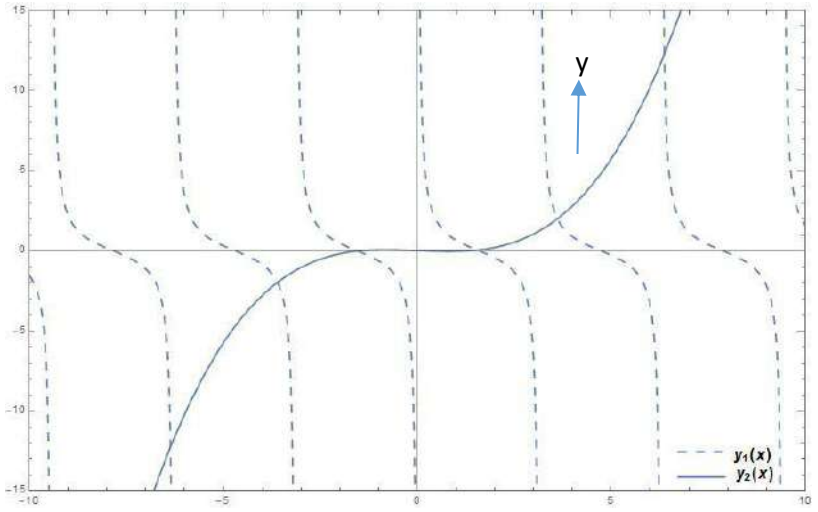


Figure 3: Nature of the imaginary roots when εK<1

It is clear from figure 3 and figure 4 that for both the cases there exists infinite number of roots. Therefore, it can be concluded that there exists infinite number of imaginary roots of the dispersion equation (3.3) that are $\pm\alpha_i$, i =1, 2, 3, 4,……. . The imaginary roots lies within the range $\alpha_1 < \alpha_2 < 2\pi < \alpha_3 < 3\pi < \alpha_4 <$ …….. . The roots also have the following property

$\alpha_n > (n-1)\pi$.

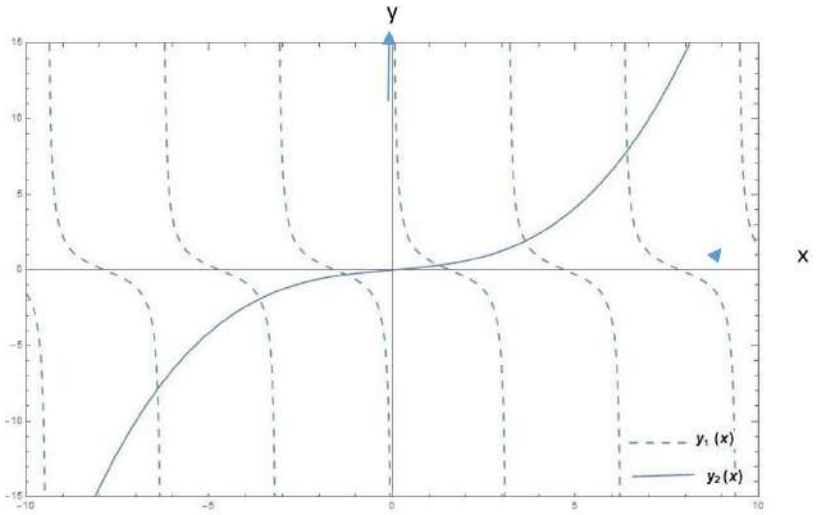


Figure 4: Nature of the imaginary roots when $\varepsilon K>1$

It can be easily shown that this dispersion equation does not have any other root by applying Rouche's theorem.

References

1. Peters, A. S. (1950). The effect of a floating mat on water waves. Commun. Pure Appl. Math. 3, 319-354.
2. Rhodes-Robinson, P. (1982). Note on the effect of surface tension on water waves at an inertial surface. *J. Fluid Mech.*, *125*, 375-377.
3. Rhodes-Robinson, P. F. (1984). On the generation of water waves at an inertial surface. *ANZIAM J*, *25*, 366-383.
4. Wehausen, J. V., & Laitone, E. V. (1960). Surface waves. In *Fluid Dynamics/Strömungsmechanik* (pp. 446-778). Berlin, Heidelberg: Springer Berlin Heidelberg.
5. Weitz, M. & Keller, J. B. (1950). Reflection of water waves from floating ice in water of finite depth. Commun. Pure Appl. Math. 3, 305-318

Chapter - 9

A Study on Internal Wave Velocity in a Two-Layer Fluid Medium in Context of the Solution of Dispersion Equation

Author

Arijit Das
Department of Applied Mathematics, Swami Vivekananda University, Kolkata, West Bengal, India

Chapter - 9

A Study on Internal Wave Velocity in a Two-Layer Fluid Medium in Context of the Solution of Dispersion Equation

Arijit Das

Abstract

This note addresses the determination of the characteristics of the roots of a dispersion equation. This equation emerges during the examination of small-amplitude internal waves in two separate, non-mixing fluids. In this scenario, the upper fluid possesses a free surface, while the lower fluid has a solid bottom. The dispersion equation's roots are identified by analyzing the graphs of function. The effect of density and width of the two fluids on the roots of the dispersion equation have been studied numerically and conclusion made the change in wave velocity.

Keywords: Dispersion equation, progressive waves. wave velocity, surface wave, interface wave.

Introduction

In the realm of fluid dynamics, the dispersion of water waves typically pertains to frequency dispersion, where waves of varying wavelengths move at distinct phase velocities. These water waves, within this context, represent waves propagating across the surface of water, with gravity and surface tension acting as the forces that restore equilibrium. Consequently, a body of water with an open surface is generally acknowledged as a medium that exhibits dispersion. In the context of a specific water depth, surface gravity waves, which occur at the interface between air and water and are restored to flatness solely by gravity, accelerate as their wavelength increases. Conversely, for a fixed wavelength, gravity waves travel at a higher phase velocity in deeper water compared to shallow water. In contrast to the behavior of gravity waves, capillary waves move at a swifter pace when they possess shorter wavelengths.

In the literature, researchers have long focused on different categories of small amplitude water waves, primarily studying them within the framework

of linearized water wave theory for several decades. These waves are often referred to as surface gravity waves because gravity serves as the sole external force acting upon them. The connection between the wave's wave number and angular frequency is termed the 'Dispersion equation.' This terminology is rooted in the observation that surface gravity waves with varying wavelengths travel at distinct speeds, resulting in a dispersion phenomenon. This dispersion arises because the wave propagation speed(c) is determined by the ratio of angular frequency (ω) to wave number (k) i.e.

$$c = \frac{\omega}{k}. \tag{1.1}$$

The basic assumptions of the linearized theory are as follows. It is presumed that water behaves as an inviscid, incompressible, and uniform fluid. The motion within it is solely influenced by gravity and is characterized as irrotational. Furthermore, the assumption regarding the smallness of motion implies that the velocity components, along with their respective derivatives, are considered to be of a first-order small magnitude.

The dispersion equation is decisive for determining the velocity potential *viz*., for an fluid of infinite depth the dispersion equation can be expressed as follows (cf. Lamb (1924))

$$k = \frac{\omega^2}{g} \ (\equiv K) \tag{1.2}$$

where g is the acceleration due to gravity. For the two dimensional case where φ depends on *x,y* only, progressive wave solution is

$$\varphi = e^{-ky \pm ikx} \tag{1.3}$$

Also there exists local solution for φ in the form

$$\varphi = e^{-k|x|}(k \cos ky - k \sin ky) \tag{1.4}$$

where k is any positive number.

Thus a general solution of complex valued potential associated with velocity potential φ*(x,y)* for infinite depth of fluid is of the form [cf, Mandal and De(2006)]

$$\varphi(x, y) = A0e^{-ky+ikx} + B0e^{-ky-ikx}$$

$$+ \quad \{\int_0^\infty A(k)e^{-i\,k\,|x|}(K\,cos\,cos\,ky - ksin\,ky)dk, x > 0 \int_0^\infty B(k)e^{i\,k\,|x|}(K\,cos\,cos\,ky - ksin\,ky)dk, x < 0$$

Where *A(k)* and *B(k)* are functions *k* only.

Again, if the fluid is of finite constant depth "*h*", a possible solution of φ for the two-dimensional case is

$$\varphi = \cosh(h - y)e^{kx}. \quad (1.5)$$

where the wave number k satisfies the dispersion equation(cf. Wehausen and Laitone(1960))

$$k \tanh kh = K. \quad (1.6)$$

This equation has two real roots, one positive, k_0 say, one negative, $-k_0$ and an infinite number of purely imaginary rots $\pm ik_n$, n= 1,2, where

$$k_n \tan k_n h + K = 0. \quad (1.7)$$

The results corresponding the finite depth of fluid is

$$\varphi(x, y) = A_0 \cosh k_0(h - y)e^{ik_0 x} + B_0 \cosh k_0(h - y)\, e^{-ik_0 x}$$

$$\begin{cases} \sum A_n e^{-k_n x} cos(h - y)), \; x > 0 \\ \quad + \\ \sum B\, e^{k_n x} cos(h - y)), x < 0 \end{cases}$$

The theory of a two-layer fluid is also gained interest of the researchers as it is more realistic in nature. There are a significant number of studies that have been conducted in this regard (cf. Linton and McIver (1995), Linton and Cadby (2001), Das and Mandal (2006), Das *et al*. (2023)). Here in the present study, the effect of different parameters on the solution of dispersion equation have been studied as it is evident from equation (1.1) that the wave number has a significant for the speed of the progressive wave. For example, a wave of smaller wave number travels faster than a wave of larger wave number.

Dispersion equation for small amplitude internal waves

The governing equations for a two-layer fluid is well known in the literature. Therefore, it will not mentioned here. Rather the primary focus is on the dispersion relation and effect of different parameters on the solution of the equation.

If the upper fluid is of uniform finite depth h above the mean interface and has a free surface while the lower fluid extends infinitely downwards, then the corresponding equation is (c.f. Linton and Mciver (1995))

$$(k - K)\{k\,(\sigma + e^{-2kh}) - (1 - e^{-2kh})\} = 0. \quad (2.1)$$

Here $\sigma=\frac{1+s}{1-s}$ where, $s = \frac{\rho_2}{\rho_1}$, $(\rho_2 < \rho_1)$, ρ_1 and ρ_2 being the densities of the lower and upper fluid respectively. The upper fluid is of uniform finite depth h above the mean interface.

This equation has two real roots, one is K and the other is say m, where m satisfies the equation

$$(\sigma + e^{-2mh}) = m(1 - e^{-2mh}), \quad (2.2)$$

So that

$$K\sigma < \mathrm{m} < K\,\frac{\sigma+1}{1-e^{-2K\sigma h}} \quad (2.3)$$

Then there exist time harmonic progressive waves with two different wave numbers *K* and *m*.

It is important to note here that, two wave numbers *K* and *m* corresponds to two different propagating waves at the free surface and the interface respectively. Also, it can be observed that the wave number of the surface wave does not depend on the density or the depth of the upper layer but the wave number corresponding to the interface wave odes depend on the various parameter that will be discussed in the next section.

If the lower fluid is not very deep then then it can be modelled as a two layer fluid wherein the lower fluid is of uniform finite depth H., say, below the mean interface, and as before the upper fluid is height h above the interface and has a free surface. In this case the dispersion relation (2.1) modifies to

$$k^2(1 - s) - kK(\coth kh + \coth kH) + K^2(s + \coth kh \coth kH) = 0. \quad (2.4)$$

This equation is given by Sherief *et al*. [6,7] while investigating forced gravity waves due to a plane and a cylindrical vertical porous wave - maker in a two layer fluid. Das and Mandal gave an detailed note about the nature of the roots of the equation (2.4). In this paper the changes in this roots will be studied for different parameters.

Solution and discussion

It can be observed from (2.1) that when the lower fluid has infinite depth the changes due to parameter change only valid for the interface wave. In fact from equation (2.2) it is observed that the lower limit for the wave number for root *m* is $K\sigma$ which increases as the value of s reaches closer to 1. That is as the density of the two fluids are not very different the waves travel very slowly. Also, if the depth of the upper fluid increases, the upper limit of *m* decreases implying the existence of faster waves.

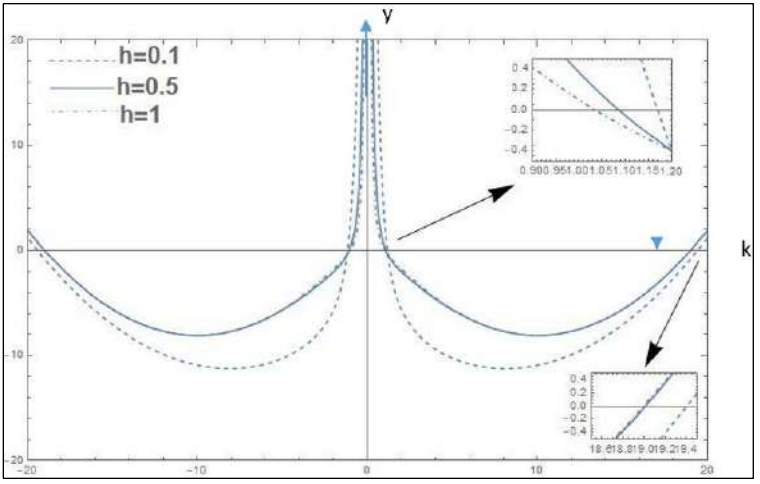


Figure 1: *f(k)* against *k* for different values of h

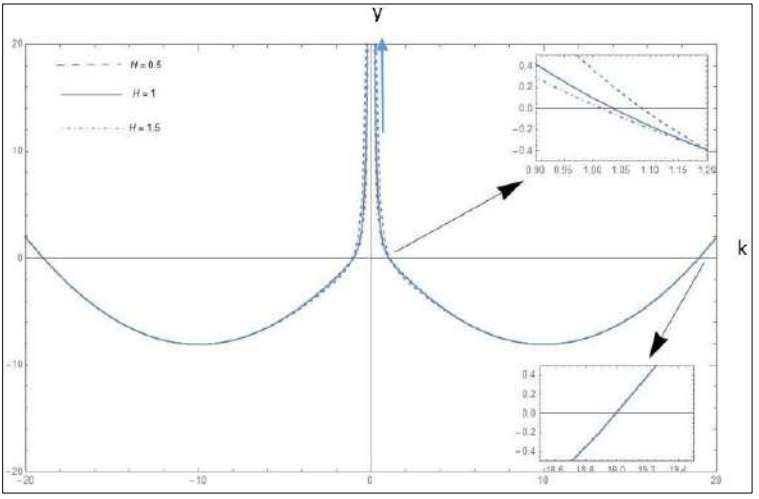


Figure 2: *f(k)* against *k* for different values of H

The roots of equation (2.4) and its changes will be studied by using graphical method.

The graphs of

$$f(k)= k^2(1-s) - kK(\coth kh + \coth kH) + K^2(s + \coth kh \coth kH)$$

have been plotted against k and the changes in the roots have been observed.

All the variables have been non-dimensionalized before plotting.

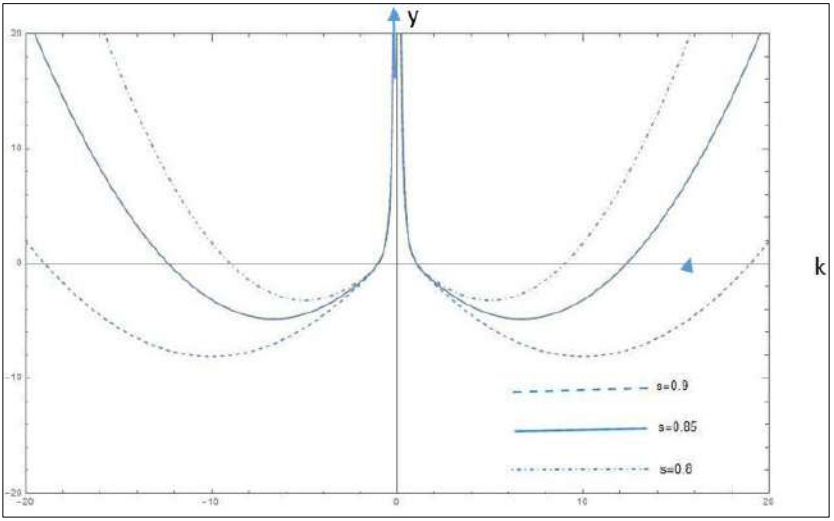


Figure 3: *f(k)* against *k* for different values of s

Figure 1 depicts the effect of different widths of the upper layer on the wave number. The wavenumber for both the surface and interface waves increases as h decreases. Therefore, the surface and interface waves travel faster if the width of the upper layer is large. In Figure 2, it can be observed how the wave number is influenced by varying widths of the lower layer also. As the parameter 'H' increases, the wave number for the surface wave also decreases but the change in the wavenumber for interface wave is not significant. Consequently, if the Lower layer's width is relatively small, the surface waves will travel at a lower speed.

In Figure 3, effect of density ratio on wavenumbers have been shown. Although the density ratio does not have such impact on the surface wave but it has huge impact on the interface wave as can be seen from figure 3. As the value of s decreases the interface wavenumber reduces significantly.

Consequently, if the lower layer is much heavier than the upper fluid then interface wave travel much faster.

References

1. Das, A., De, S., & Mandal, B. N. (2023). Wave interaction with an elliptic disc submerged in a two-layer fluid. *Applied Mathematical Modelling*, *117*, 786-801.
2. Das, D., & Mandal, B. N. (2005). A note on solution of the dispersion equation for small-amplitude internal waves. *Archives of Mechanics*, *57*(6), 493-501.
3. Lamb, H. (1924). *Hydrodynamics*. University Press.
4. Linton, C. M., & Cadby, J. R. (2002). Scattering of oblique waves in a two-layer fluid. *Journal of Fluid Mechanics*, *461*, 343-364.
5. Linton, C. M., & McIver, M. (1995). The interaction of waves with horizontal cylinders in two-layer fluids. *Journal of Fluid Mechanics*, *304*, 213-229.
6. Mandal, B. N., & De, S. (2015). *Water wave scattering*. CRC Press.
7. Wehausen, J. V., & Laitone, E. V. (1960). Surface waves. In *Fluid Dynamics /Strömungsmechanik* (pp. 446-778). Berlin, Heidelberg: Springer Berlin Heidelberg.

Chapter - 10

Formation of a Weakly Singular Integral Equation arising in the Problem of Water Wave Scattering by a Thin Vertical Barrier Submerged in Deep Water and its Numerical Solution using Bernstein Polynomial

Authors

B Sarkar
School of Basic Sciences, Swami Vivekananda University, Barrackpore, Kolkata, West Bengal, India

P Sharma
School of Basic Sciences, Swami Vivekananda University, Barrackpore, Kolkata, West Bengal, India

Chapter - 10

Formation of a Weakly Singular Integral Equation Arising in the Problem of Water Wave Scattering by a Thin Vertical Barrier Submerged in Deep Water and its Numerical Solution using Bernstein Polynomial

B Sarkar and P Sharma

Abstract

A weakly singular integral equation arising in the problem of water wave scattering by a thin vertical barrier submerged in infinitely deep water is formed and that is solved here by employing a numerical method based on polynomial approximation based on Bernstein polynomials as basic functions. The reflection and transmission coefficients for the scattering problem are then obtained numerically using the numerical solution of the integral equation. The numerical results obtained by the present method are in excellent agreement with those obtained by different methods in the literature. The same method also been successfully employed to solve the integral equation arising in the complementary problem of water wave scattering by a thin vertical barrier partially immersed in deep water and the corresponding reflection and transmission coefficients are obtained numerically whose agreement with known results is excellent.

Keywords: Weakly singular integral equation, wave scattering, vertical barrier, Bernstein polynomials, reflection and transmission coefficients.

Introduction

The classical problem of water wave scattering by thin vertical barrier submerged in deep water and extending infinitely downwards leads to the weakly singular integral equation (cf. Mandal and Chakrabarti [1])

$$\int_0^b f(u)\mathrm{K}(y, u)\,du = e^{-ky},\ 0<y<b\ldots \quad (1.1)$$

where

$$K(y,u) = \frac{1}{pi}\left[K\ ln\ ln\ \left|\frac{y+u}{y-u}\right| - 2e^{-K(y+u)}\int_{-\infty}^{K(y+u)} \frac{e^v}{v}dv\right] + 2ie^{-K(y+u)}, 0 < y,u < b \quad (1.2)$$

'b' being the depth from which the barrier is submerged below the mean free surface of water, $K(>0)$ being the wave number, and $f(y)$is the horizontal component of velocity at a depth y in the gap above the barrier. The reflection and transmission coefficients R, T are given by

$$T = 1 - R = -2i\int_0^b f(u)\, e^{-Ku}du \quad (1.3)$$

The explicit solution of the integral equation (1.1) was obtained in the literature by reducing it to a strongly singular first kind integral equation with Cauchy type kernel whose solution essentially depends on the behavior of $f(y)$ at the end points $y = 0(f(0) \neq 0)$ and $y = b\left(f(y) = 0(b-y)^{-\frac{1}{2}} as\ y \to b-0\right)$. The reflection and transmission coefficients were then obtained in closed form in terms of modified Bessel functions. However, the method appears to be somewhat complicated. Instead of reducing the weakly singular integral equation (1.1) to a first kind strongly singular integral equation with Cauchy type kernel, one can straightway solve (1.1) numerically and use this numerical solution to compute $|R|$ and $|T|$ numerically using the relation (1.3). This procedure is adopted here. The integral equation (1.1) defined over (0, b) is transformed to a similar type of integral equation defined in (0,1) to facilitate the use of Bernstein polynomials in obtaining the approximate numerical solution. In fact, the Bernstein polynomials have been used to solve some linear and nonlinear ordinary and partial differential equations approximately by Bhatta and Bhatti [2] and Bhatti and Bracken [3] and some integral equations by Mandal and Bhattacharya ([4), [5]). Recently, Gayen *et al.* [6] solved the integral equation somewhat similar to (1.1) which arises for the complementary problem of water wave scattering by a thin vertical barrier partially immersed up to depth a below the mean free surface in deep water, approximately by using simple polynomials and also Chebyshev polynomials in approximation of the unknown function, after changing the range of integral equation from (a, ∞) to (0, 1). This integral equation has also been solved here using Bernstein polynomials after changing the range from (a, ∞) to (0, 1). The reflection and transmission coefficients for the both the scattering problems are obtained numerically and are depicted

graphically against the wave number Kb for submerged barrier problem and Ka for the partially immersed problem in figures 1 and 2. The curves in these figures are seen to coincide with the curves given in Ursell [7]. This demonstrates the correctness of the method employed here.

Formation of weakly singular integral equation

For two-dimensional motion, the solutions for the water wave potential $\emptyset(x, y)$ can be obtained by employing the method of separation of variables for solving the two dimensional Laplace equation $\nabla^2\emptyset = 0$ in the fluid region (2.1),

Where $\nabla^2 = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}$

Along with free surface and bottom conditions,

$$K\emptyset + \frac{\partial \emptyset}{\partial y} = 0 \; on \; y = 0$$

Where $K = \frac{\sigma^2}{g}, \sigma = angular \; frequency$

$$\nabla\phi \to 0 \; as \; y \to \infty$$

For deep water, the progressive wave solution is given by

$$\emptyset(x, y) = \; e^{-Ky \pm iKx} \tag{2.2}$$

and local solutions are

$$\emptyset(x, y) = \; (kcos\; ky \; - \; Ksin\; ky)e^{-k|x|} \tag{2.3}$$

where k is any real and positive number.

For two-dimensional water wave problems, the general solution for the potential function $\emptyset(x, y)$ can be constructed by a linear superposition of appropriate progressive wave solutions and the local solution.

In case of deep water, the general solution for $\emptyset(x, y)$ in their region $x > 0$ is given by

$$\phi(x, y) = ce^{-ky+ikx} + de^{ky-ikx} + \int_0^\infty \; A(k)(kcosky - ksinky)e^{-kx}dk$$

where c, d are constants and $A(k)$ is a function of k.

[by Havelock expansion of two-dimensional water wave potential for deep water in region $x > 0$]

1. Reduction to a singular integral equation

We have

$$\nabla^2\phi = 0 \; in \; y \geq 0 \tag{3.1}$$

$$K\phi + \frac{\partial \phi}{\partial x} = 0 \; on \; y = 0 \tag{3.2}$$

$$\frac{\partial \phi}{\partial x} = 0 \; on \; x = 0, y \; \in (b, \infty) \tag{3.3}$$

$$r^{-\frac{1}{2}} \nabla\phi \; is \; bounded \; as \; r \to 0 \tag{3.4}$$

where r denotes the distance from the submerged edge or edges of the barriers.

$$\nabla\phi \to 0 \; as \; y \to \infty \tag{3.5}$$

$$\phi \sim \phi^{inc}(x, y) + R\phi^{inc}(-x, y) \; as \; x \to -\infty$$

$$T\phi^{inc}(x, y) \; as \; x \to \infty \tag{3.6}$$

A suitable representation of $\emptyset(x, y)$ satisfying the equations (3.1) and conditions (3.2), (3.5), (3.6) is obtained by utilizing Havelock expansion in the form given by,

$$\phi(x, y) = T\phi^{inc}(x, y) + \int_0^\infty A(k)(kcosky - ksinky)e^{-kx}dk, x > 0$$

$$= \phi^{inc}(x, y) + R\phi^{inc}(-x, y) + \int_0^\infty B(k)(kcosky - ksinky)e^{-kx}dk, x < 0 \tag{3.7}$$

where $A(k), B(k)$ are unknown functions of k.

Let $\frac{\partial \phi}{\partial x}(0, y) = f(y), y > 0$ (3.8)

Then, $f(y)$ denotes the horizontal component of velocity in the plane of barrier and thus

$$f(y) = 0, y \; \in (b, \infty) \tag{3.9}$$

[y-axis is chosen vertically downwards into the water and the plane $y = 0$ denotes the rest position of free surface and let a thin vertical barrier occupy the position $x = 0, y \; \epsilon(b, \infty)$]

And because of the edge condition

$$f(y) = 0(|c - y|^{-\frac{1}{2}} \; as \; y \to c \tag{3.10}$$

where c denotes the depth of an edge of the barrier beneath the mean free surface. Using the representation (3.7) in the relation (3.8) for $x = 0^+$ and $x = 0^-$, we find that

$\phi^{inc}(x,y) = e^{-ky+ikx}$ and

$$ikTe^{-ky} - \int_0^b kA(k)(kcosky - ksinky)dk = ik(1-R)e^{-ky} + \int_0^b kB(k)(kcosky - ksinky)dk$$

$$= f(y), y > 0 \quad (3.11)$$

Using Havelock expansion formula,

$$f(y) = 2ke^{-ky}\left(\int_0^b f(u)e^{-ku}du\right) + \frac{2}{\pi}\left(\int_0^b f(u)\frac{(kcosku-ksinku)}{k^2+u^2}\,du\right) \quad (3.12)$$

Then by comparison,

$$T = 1 - R = 2i\int_0^b f(u)e^{-ku}du \text{ and} \quad (3.13)$$

$$A(k) = -B(k) = -\frac{2}{\pi}\left(\int_0^b f(u)\frac{(kcosku-ksinku)}{k(k^2+u^2)}\,du\right) \quad (3.14)$$

Then, the unknown functions $A(k), B(k)$ and unknown constants are expressed in terms of integrals involving the single unknown function $f(y)$. We now obtain an integral equation for $f(y)$.

Since, the potential function $\emptyset(x,y)$ must be continues across the gap in the plane of vertical barrier we must have, $\emptyset(+0,y) - \emptyset(-0,y) = 0, y\epsilon(0,b),$. (3.15)

Using the representation (3.7) in equation (3.15) and utilizing the relation, we obtain an integral equation for $f(y)$ [after changing the order of integration which we assume to be valid] in the form

$$\frac{2}{\pi}\left(\int_0^b f(u)\int_0^\infty \frac{(kcosku-ksinku)(kcosky-ksinky)}{k(k^2+u^2)}\,dk\,du\right) = -Re^{-ky}, 0 < y < b \quad (3.16)$$

Which implies (cf. F. Ursell [7])

$$\int_0^b f(u)\left(\frac{1}{pi}\left[K\ ln\ ln\left|\frac{y+u}{y-u}\right| - 2e^{-K(y+u)}\int_{-\infty}^{K(y+u)}\frac{e^v}{v}dv\right] + 2ie^{-K(y+u)}\right)du = e^{-ky}, 0 < y < b \ldots \quad (3.17)$$

Bernstein polynomial method

Introduction

The Bernstein polynomials of degree n over the interval $[a,b]$ is defined by

$$B_{i,n}(x) = \binom{n}{i} \frac{(x-a)^i (b-x)^{n-i}}{(b-a)^n}, i = 0,1,2, \ldots, n. \quad (4.1)$$

These Bernstein polynomials were first used by Bernstein in the proof of Weirstrass theorem by a linear substitution the interval $[a, b]$ can be transformed into $[0,1]$ so that without any loss of generality one can consider a function $f(x)$ to be defined on $[0,1]$. The corresponding representation the function $f(x)$ on the Bernstein polynomial basis in the interval $[0, 1]$ is given by,

$$B_n^f(x) = \sum_{i=0}^{n} \quad a_i \binom{n}{i} x^i (1-x)^{n-i}, \quad (4.2)$$

a_i for the known function $f(x)$ is defined by $a_i = f\left(\frac{i}{n}\right)$

(4.2) is the Bernstein Representation of the function $f(x)$ and it can be proved that for a function $f(x)$ bounded on $[0, 1]$ the relation.

$$B_n^f(x) = f(x) \quad (4.3)$$

holds at each point of continuity x of $f(x)$ and that the relation holds uniformly on $[0,1]$ if $f(x)$ is continuous on the interval. In fact, by suitable changes in the definition of coefficient a_i in the expression (2.2) the above property can be extended to approximately any general function. The coefficients a_i in which case can be an arbitrary constant.

Hence Bernstein polynomial basis provides a very good basis for approximation of unknown function in differential or an integral equation.

The following figures give a diagrammatic representation of Bernstein polynomial $Bi, n(x), i = 0,1,2,3$. Bernstein polynomial of degree 1 is given by

$$B_{0,1} = (1-x), B_{1,1} = x$$

Bernstein polynomial of degree 2 is given by,

$$B_{0,2}(x) = (1-x)^2, B_{1,2}(x) = 2x(1-x), B_{2,2}(x) = x^2$$

Bernstein polynomial of degree 3 are given by

$$B_{0,3}(x) = (1-x)^3, \; B_{1,3}(x) = 3x(1-x)^2, B_{2,3}(x) = 3x^2(1-x), B_{3,3}(x) = x^3$$

Bernstein polynomial of degree 4 are given by

$$B_{0,4}(x) = (1-x)^4, B_{1,4}(x) = 4x(1-x)^3, B_{2,4}(x) = 6x^2(1-x)^2$$

$$B_{3,4}(x) = 4x^3(1-x), B_{4,4}(x) = x^4$$

Properties of bernstein polynomials

Recursive relation

Any i[th] Bernstein polynomial of degree n can be expressed by a linear combination Bernstein polynomials of degree (n-1), i.e. for x∈ [0,1]

$$B_{i,n}(x) = (1-x)\, B_{i,n-1}(x) + x\, B_{i-1,n-1}(x)$$

Derivatives

$$\frac{d}{dx} B_{i,n}(x) = \frac{d}{dx}\left(\binom{n}{i} x^i (1-x)^{n-i}\right)$$

$$= \binom{n}{i}\left(i x^{i-1}(1-x)^{n-i} - (n-i)x^i(1-x)^{n-i-1}\right)$$

$$= n\left(\binom{n-1}{i-1} x^{i-1}(1-x)^{n-i} - \binom{n-1}{i} x^i (1-x)^{n-i-1}\right)$$

$$= B_{i-1,n-1}(x) - B_{i,n-1}(x)$$

Positivity

On the interval [0,1] it follows from definition of Bernstein polynomial that for all x∈ [0,1], they are always positive.

Symmetry

It follows from the definition of Bernstein polynomials that

$$B_{i,n}(x) = B_{n-i,n}(x)$$

Normalization

The Bernstein polynomials of any degree always sum to one as shown below. Using the recursive relation for $B_{i,n}(x)$, we have

$$\sum_{i=0}^{n} B_{i,n}(x) = \sum_{i=0}^{n} \left[(1-x)\, B_{i,n-1}(x) + x\, B_{i-1,n-1}(x)\right]$$

$$= (1-x)\sum_{i=0}^{n} B_{i,n-1}(x) + x\sum_{i=1}^{n} B_{i-1,n-1}(x) = \sum_{i=0}^{n} B_{i,n-1}(x) = (1-x) + x = 1$$

Therefore,

$$\sum_{i=0}^{n} B_{i,n}(x) = \sum_{i=0}^{n-1} B_{i,n-1}(x) = \sum_{i=0}^{n-2} B_{i,n-2}(x) = \cdots \sum_{i=0}^{1} B_{i,1}(x)$$

Degree raising

Any Bernstein polynomial of degree less than n can be expressed as a linear combination of Bernstein polynomials of degree n

$$B_{i,n-1}(x) = \frac{n-1}{n} B_{i,n}(x) + \frac{i+1}{n} B_{i+1,n}(x)$$

These are some of the properties of the Bernstein polynomial.

Now, put $u = bv, y = bz$ to reduce (3.17) to the form

$$\int_0^1 \quad bf(bv)k(bz, bv)dv = -e^{kbz}, 0 < z < 1 \tag{5.1}$$

To find an approximate solution of (2.1), $f(b, v)$ approximated by using Bernstein polynomials in $[0,1]$ as

$$\sum_{i=1}^{n} \quad a_i B_{i,n}(v) = bf(bv) \tag{5.2}$$

where a_i(i=0,1,2,3……n) are unknown constants to be determined.

Substituting (2.2) in (2.1) we get

$$\sum_{i=1}^{n} \quad a_j A(Z) = e^{-kbz} \tag{5.3}$$

$$A(z) = \int_0^1 \quad B_{j,n}(v)k(bz, bv)dv$$

To find the $(n+1)$ constants a_i we collocate at $(n+1)$ points $z = zi(i = 0,1,2 \ \dots \ \dots \ n)$ where $Z_i = \frac{i}{n}, (i = 0,1,2 \dots, n) so\ that\ 0 \le z_i \le 1$.

Reflection and transmission coefficients

Substituting the expression for $f(u)$ from (1.1) in (1.3), we obtain

$$R = 1 + 2i \sum_{j=1}^{n} \quad a_j \int_0^1 \quad B_{j,n}(v)e^{-kbv}dv \tag{6.1}$$

Choosing a suitable n, we solve the linear system of algebraic equation (5.3) and obtain $aj, (j = 0,1,2, \dots, n)$, Using these values of a_j, we find R from (5.1). T is obtained from the relation $|T|^2 = 1 - |R|^2$. Numerical estimates for the actual reflection and transmission coefficients $|R|$ and $|T|$ are now obvious.

Numerical results

The numerical computations of the reflection coefficients as given by (3.1) for completely submerged barrier produce almost same numerical

results calculated from the exact result given by Ursell [7] and are displayed in table 1.

Table 1: $|R|$ for completely submerged barrier (n=100)

Kb	0.1	0.5	0.9	1.3	1.7
Present method	0.610402	0.266663	0.126674	0.0601705	0.0288083
Ursell [7]	0.610417	0.266672	0.1266696	0.060183	0.0288162

Using a similar method, the reflection coefficient for a partially immersed barrier has been obtained but the details are not given. It is observed that the numerical results are obtained by the present method and obtained from Ursell's [7] exact result coincides up to 4 decimal places. These are displayed in Table 2.

Table 2: $|R|$ for partially immersed barrier (n= 100)

Ka	**0.1**	**0.5**	**0.9**	**1.3**	**1.7**
Present Method	0.0159439	0.439360	0.908946	0.989142	0.998472
Ursell [7]	0.0159589	0.439378	0.908905	0.98912	0.998452

$|R|$, $|T|$ computed by the present method for both submerged and partially immersed barrier are depicted graphically and compared with the curves given in Ursell [7]. It is observed that the numerical results, obtained by the present method, are found to be in excellent agreement with Ursell's [7] result.

Conclusion

A simple method of approximating unknown function in terms of truncated series involving Bernstein polynomials is proposed here to solve a first kind integral equation which arises in the problem of water wave scattering by a barrier. The reflection and transmission coefficients are obtained numerically. Excellent agreement is seen to have been achieved between the approximate and exact solutions. This method appears to be straight forward and avoids the use of complicated expression involving Bessel functions.

References

1. B.N. Mandal. A. Chakrabarti, Water wave scattering by barriers, WIT Press/ Computational Mechanics, 2000.
2. D.D. Bhatta, M.L. Bhatti, Numerical solution of Kdv equation using Bernstein polynomials, Appl. Math. Comput. 174,(2006) 1255-1268.

3. M. I. Bhatta, P. Bracken, Solution of Differential equation in a Bernstein polynomial basis, J. Comput. Appl. Math. 205,(2007), 272-280.
4. B. N. Mandal, S. Bhattecharye, Numerical solution of some classes of integral equations using Bernstein polynomials, Appl. Math and Computation. 190 (2007) 1707-1716.
5. S. Bhattecharye, B.N. Mandal, Use of Bernstein polynomials in numerical solutions of Volterra integral equations, Appl. Mathematical Sci. 2 (2008) 1773- 1787.
6. R. Gayen, S. Gupta, A. Chakarbarti, Approximate solution of the problem of scattering of surface water waves by a partially immersed rigid plane vertical barrier, Appl. Math. Lett. 58 (2016) 19-25.
7. F. Ursell, The effect of a fixed vertical barrier on surface waves in deep water, Math. Proc. Camb. Phil. Soc. 43(3) (1947) 374-382.

Chapter - 11

Revisited Study on Advection Dominated Accretion Flow Around Black Hole

Authors

P Sharma
School of Basic Sciences, Swami Vivekananda University, Barrackpore, Kolkata, West Bengal, India

Subhabrata Mondal
School of Basic Sciences, Swami Vivekananda University, Barrackpore, Kolkata, West Bengal, India

Chapter - 11

Revisited Study on Advection Dominated Accretion Flow Around Black Hole

P Sharma and Subhabrata Mondal

Abstract

In order to explain the low efficiency by accretion in a binary system and accounts for flows with energy advection, ADAF model was proposed. The physical picture being this: during the accretion onto black holes, the gas is "forced" either not to radiate its gravitational or to direct the radiated flux straight to a black hole, all energy is absorbed together with the matter. It can be realized when the accretion rate is very small. To explain this model for low luminous system several analytical solutions came up. The widely popular solutions like self-similar and global solution for ADAF is explained.

Keywords: Accretion flow, accretion disk, black hole physics, galaxies: active nuclei, advection.

Introduction

The process by which compact stars gravitationally capture ambient matter is called Accretion. An accretion disk is a structure formed by diffused material in orbital motion around massive central body (typically a star). Friction causes orbiting material in the disk to spiral inward towards the central body. Gravitational and frictional forces compress and raise the temperature of the material, causing emission of electromagnetic radiation.

Possible modes of accretion onto a black hole

- Via an accretion disk (geometrically thin, optically thick)
- Via a hot flow (more spherically extended, optically thin) advection-dominated accretion flow (ADAF) due to the nature of the very hot gas the flow cannot radiate away the energy released when the gas settles in the gravitational potential.

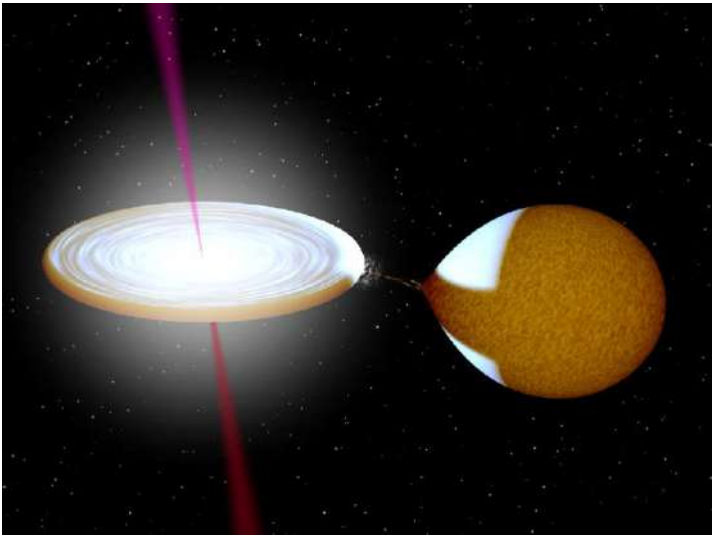


Figure 1: Schematic diagram of accretion around black hole

To explain the electromagnetic spectrum emitting from an accretion disk various types of models comes up. In this manuscript, we will mainly discuss about the ADAF model around black hole. When the accretion rate is sub-Eddington and the opacity very low, an ADAF is formed. Advection-dominated accretion flows (ADAFs) are important and also one of the fundamental solutions for the accretion flows into the black holes accretion in low-luminous systems. The ADAF solution has received much attention in recent years because it successfully explains why some nearby galaxies are so dim even though their accretion rates are not very small. This type of accretion disk was predicted in 1977 by Ichimaru. Some elements of the ADAF model were present in the influential 1982 ion-tori paper by Rees, Phinney, Begelman and Blandford. ADAFs started to be intensely studied by many authors only after their rediscovery in the mid-1990 by Narayan and Yi, and independently by Abramowicz, Chen, Kato, Lasota (who coined the name ADAF), and Regev. Most important contributions to astrophysical applications of ADAFs have been made by Narayan and his collaborators.

The ADAF solution has received much attention in recent years because it successfully explains why some nearby galaxies are so dim even though their accretion rates are not very small. The best evidence of ADAF comes from the supermassive black hole in our Galactic centre, Sgr A*. In past studies, several methods have been used to obtain the transonic

solutions of ADAFs, e.g., the relaxation scheme (e.g., Narayan *et al*. 1997), the implicit integration scheme (e.g., Kato *et al*. 1998), the outward integration scheme with the inner boundary condition (Becker & Le 2003), or the inward or outward integration scheme with the special condition at and inside the sonic radius (Lu *et al*. 1999). All these works assume the diffusive viscosity with an infinite diffusive velocity. A prominent feature of an ADAF compared to the standard thin disk is its low radiative efficient at low accretion rates.

An advection dominated accretion flow is one in which most of the energy is released by viscous dissipation is stored in the gas and advected in the centre, and only a small fraction of energy is radiated. Advection dominated accretion can occur in two different limits:

1) At very high mass accretion rates, radiation trapped in the accreting gas because of the large optical depth and it is advected with the flow. This typically occur for $\dot{M}>\dot{M}_{edd}$
2) At sufficiently low accretion rate $\dot{M}$, the accreting gas become optically thin. The cooling time of the gas is then longer than the accretion time, and once again we have an ADAF.

This low $\dot{M}$ brach of ADAF has several appealing features for explaining the observation of Sgr A*. The Narayan *et al*. (1995) ADAF model of Sgr A* gave a reasonable fit to the observations available at that time and provide an explanation for the low luminosity of the source.

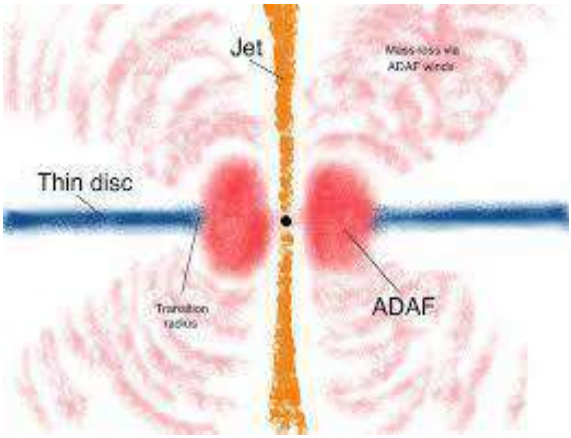


Figure 2: A cartoon version of the most favoured scenario for the black hole accretion flow geometry and outflows in low-luminosity AGNs

Basic need of ADAF model

- Standard thin disk accretion model (Shakura & Sunyaev 1973; Novikov & Throne 1973) is a great model for some sources but for other sources and even in same sources at different times are not consistent with the thin disk model.
- If we want to go beyond empherical data-fitting,we need additional physical model.
- To explain low luminosity systems
- To explain the Sgr A* system which is extremely under luminous.

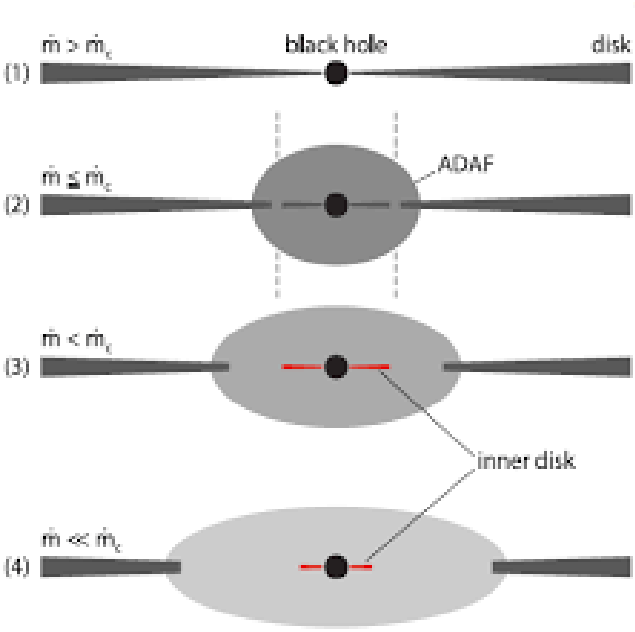


Figure 3: Schematic diagram of ADAF around black hole with different critical mass accretion rate

Basic equations

The accretion flow onto a black hole is governed by 4 basic equations:

1) Conservation of mass,
2) Radial momentum,
3) Angular momentum and
4) Energy equation.

(Radial and angular momentum equations are the R and ϕ - components of the angular momentum equation in cylindrical coordinates). The gravitational field is assumed to be dominated by a central point mass.

Mass conservation

$$\frac{d}{dR}\,(2\pi R)(2\rho H v) = 0$$

Integrating we get,

$$-4\pi R\rho H v = \dot{M} = \text{constant} \qquad (1)$$

Where, H =vertical half thickness of the flow, ρ=density of the accreted gas, υ is the radial velocity which is defined to be negative when the flow is inward, and $\dot{M}$ is the mass accretion rate.

Radial momentum equation (R)

$$v\frac{dv}{dR} = \Omega^2\, R - \Omega_k^2 R - \frac{1}{\rho}\frac{d}{dR}(\rho C_s^2) \qquad (2)$$

where Ω is the angular velocity of the gas.

The angular momentum equation

$$v\frac{d}{dR}(\Omega R^2) = \frac{1}{\rho R H}\frac{d}{dR}\left(\nu\rho R^3 H \frac{d\Omega}{dR}\right) \qquad (3)$$

where ν is the kinematic coefficient of viscosity. In the spirit of Shakura & Sunyaev (1973), ν can be written as, $\nu = \frac{C_s^2}{\Omega_k}$ then,

$$\frac{d\Omega}{dR} = \frac{v\Omega_k(\Omega R^2 - j)}{\alpha R^2 C_s^2}$$

α is assumed to be a constant and the integration constant j represents the specific angular momentum per unit mass accreted by the black hole.

Energy equation

$$\frac{\rho v}{(\gamma-1)}\frac{dC_s^2}{dR} - C_s^2 v\frac{d\rho}{dR} = \frac{\alpha\rho C_s^2 R^2}{\Omega_k}\frac{d\Omega}{dR}$$

$$\text{Or, } \rho\nu T\frac{dS}{dR} = q^+ - q^- = \rho\nu R^2\left(\frac{d\Omega}{dR}\right)^2 - q^- = \rho\nu R^2 f\left(\frac{d\Omega}{dR}\right)^2 \qquad (4)$$

Where, $\nu = \alpha c_s H = \alpha\frac{C_s^2}{\Omega_k}$, Where S is specific entropy, q^+ = energy generated by the viscosity per unit volume and q^-= radiative cooling per unit volume. Accreting gas is heated by the viscosity (q^+)and cooled by radiation (q^-). Any excess heat is stored in the gas and transported with the flow. This represents advection of energy (q^{adv}) or advective cooling, where γ is the

ratio of specific heats of the gas. It is useful to rewrite energy equation (5) completely as, $q^{adv} = q^{+} - q^{-}$.

Regimes of energy equation

We can identify three different regimes according to the relative importance of q:

- $q^{+} \simeq q^{-} \gg q^{adv}$; Accretion flow is cooling dominated and there is basically no advection. Example: standard thin disks (Shakura & Sunyaev 1973). Another solution: SLE (1976)-geometrically thick, optically thin, two temperature disks (the disk is puffs up because radiation is inefficient but this solution is unstable).
- $q^{adv} \simeq q^{+} \gg q^{-}$; this is called an advection dominated accretion flow (ADAF) since the viscous energy is stored in the gas and advected into the black hole (NY 1994). This also means f~1.
- $-q^{adv} \simeq q^{-} \gg q^{+}$; This corresponds to a flow where energy generation through a viscous heating is negligible. Example: Bondi-Hoyle (spherical) accretion.

To understand the Basic Properties of ADAFs. A simple analytical solution would be helpful for understanding the basic qualitative properties of ADAFs. Fortunately, such type of solutions exists.

1. Solution Height-Integrated ADAF Equations
2. Self-Similar solution
3. Simplified global solution
4. Global transonic solution

Height integrated ADAF equation

As In the standard theory of thin accretion disks, vertically average the flow equations and consider a 2D flow in the $R\phi$ plane. Also assume a steady axisymmetric flow so that, $\frac{\partial}{\partial t} = \frac{\partial}{\partial \phi} = 0$, and all flow variables are function of R. Assuming gas to pressure dominated, the pressure can be written as, P=ρC_s^2, where $\rho(R)$=height averaged density

H=$\frac{C_S}{\Omega_K}$, v_k=$\Omega_K R = \left(\frac{GM}{R}\right)^{1/2}$

$TdS = dU + pdV$, $v = \alpha \frac{c_s^2}{\Omega_K}$

If Ω_K is thekeplerian angular velocity which takesthe form,

$$\Omega_K{}^2 = \frac{GM}{(R - R_g)^2 R}$$

If the gravitational potential of the central black hole is assumed to be described by Paczynski & Wiita potential, ϕ(R) = $\frac{-GM}{(R-R_g)}$, with M being the mass of the black hole, $R_g$is the gravitational radius ≡ $\frac{2GM}{C^2}$.

Self-similar ADAF solution

Although the ADAF equations look complicated they have a simple self-similar solution in which all quantities vary as power loss of radius. Analytical approximation to the structure of optically thick and thin ADAF have been derived by NY1994. Assuming Newtonian gravity ($\Omega_k^2 = \frac{GM}{R^3}$) and taking f to be independent of R from equation (4). The equations for self-similar solution can be written as,

v(R)=$\frac{(5+2\varepsilon)}{3\alpha^2}$g(α, Ɛ)α$v_{eff}$ (5)

and Ω(R)=[$\frac{2\varepsilon(5+2\varepsilon)}{9\alpha^2} g(\alpha,\varepsilon)$]$^{\frac{1}{2}}\frac{v_{eff}}{R}$ (6)

C_s^2(R)= $\frac{2\varepsilon(5+2\varepsilon)}{9\alpha^2} g(\alpha,\varepsilon)\ v\_eff$ (7)

Where $v_{eff} = \left(\frac{GM}{R}\right)^{\frac{1}{2}}$, Ɛ=Ɛ/$f$ =$\frac{1}{f}\frac{(\frac{5}{3}-\gamma)}{(\gamma-1)}$, g(α, Ɛ)=$\left[1 + \frac{18\alpha^2}{(5+2\varepsilon)^2}\right]^{1/2} - 1$

γ is the ratio of specific heat of the gas. Setting *f=1* and taking the limit $\alpha^2 \ll 1$ the equation 5 & 7 takes the form,

$$\frac{v}{v_{ff}} \simeq -\left(\frac{\gamma-1}{\gamma-5/9}\right)\alpha, \frac{\Omega}{\Omega_k} \simeq \left[\frac{2(\frac{5}{3}-\gamma)}{3(\gamma-\frac{5}{9})}\right]^{1/2}, \frac{c_s^2}{v_{ff}^2} \simeq \frac{2}{3}\frac{(\gamma-1)}{(\gamma-\frac{5}{9})} \quad (8)$$

It appears that α≃0.1; typically, α is 0.2-0.3. This means that radial velocity of the gas in an ADAF comparable to the free fall velocity $v \gtrsim 0.1 v_{ff}$. The gas accretes quite rapidly.

The gas rotates with a sub keplerian angular velocity and is only supported by centrifugal forces. In extreme case when $\gamma \rightarrow 5/3$ the flow has no rotation. Since most of the viscously generated energy stored in the gas as internal energy, rather than being radiated, the gas temperature is quite high in fact optically thin ADAF have almost virial temperature. This causes the

gas to "puff up".

$$v_R \sim -\frac{9\alpha(\gamma-1)}{(9\gamma-5)} v_k \sim\sim -0.05 v_k \tag{9}$$

$$\Omega \simeq \left[\frac{2\left(\frac{5}{3}-\gamma\right)}{3\left(\gamma-\frac{5}{9}\right)}\right]^{\frac{1}{2}} \Omega_k \simeq 0.6\Omega_k \tag{10}$$

$$c_s \simeq \left[\frac{6(\gamma-1)}{(9\gamma-5)}\right]^{\frac{1}{2}} v_k \simeq 0.6 v_k \tag{11}$$

$$\rho \simeq \rho_0 \left(\frac{R_0}{R}\right)^{3/2} \tag{12}$$

Basic need of simplified global solution

The self-similar solution may provide an excellent description of an advection-dominated flow far from the flow boundaries as it was shown by Narayan & Yi (1994) in the case of a settling subsonic solution (i.e., accretion on to a non-magnetized star). However, the structure of the flow in the inner and outer regions may have an important influence on the general properties of the modelled objects: high energy radiation is emitted primarily close to the black hole whereas optical and UV radiation will originate mostly in the outer regions where the flow may form a Keplerian disk. The global solution rather than the self-similar solution of the ADAF equations is required. This is because most of the radiation of an ADAF comes from its innermost region where the self-similar solution breaks down. However, it is technically very difficult to calculate the global solution of an ADAF. An ADAF is transonic, and thus its global solution should satisfy the sonic-point condition in addition to the outer boundary condition. But the global solution of ADAFs is difficult to calculate because it is mathematically a two-point boundary value problem. This hampers wide application of the ADAF model. So, a new solution is proposed a simplified global solution to overcome this difficulty. Prompted by the self-similar solution of ADAFs, by replacing the radial momentum equation, which is the most difficult to handle, with a simple algebraic relation and then solve the remaining two differential equations also adjusting the two free parameters that the best approximation compared to the exact global ADAF solution can be obtained. So, a simplified global ADAF solution needed so adopting a simple algebraic relation to replace the radial momentum equation, thus avoiding the two-point boundary value problem.

Simplified global solution

The basic equations of an ADAF describe the conservation of mass, radial and azimuthal components of the momentum, and energy can be written as,

Taking equation 1 & 2 and equation 3 & 4 becomes,

$$v(\Omega r^2 - j) = -rc_s^2\alpha \tag{13}$$

$$\rho v T_i \frac{dS_i}{dR} = (1-\delta)q^+ - q_{ie} \tag{14}$$

$$\rho v T_e \frac{dS_e}{dR} = (-\delta)q^+ - q^- + q_{ie} \tag{15}$$

This simplified solution does not include outflows from the ADAF, but it is easy to extend the calculation to that case by simply using a radius-dependent mass accretion rate,

$$\dot{M} = \dot{M}_0 \left(\frac{r}{r_{out}}\right)^s$$

With s>0 being a constant. Where δ in equations 14 and 15 describes the fraction of the turbulent dissipation rate q^+ that directly heats electrons; setting δ=0.3. The quantity q_{ie} describes the energy transfer rate from ions to electrons by Coulomb collision, and q^- is the radiative cooling rate. Considering synchrotron and bremsstrahlung emissions and their Comptonization. The details of the calculation of the spectrum can be found in Yuan *et al*. Also assume a Schwarzschild black hole and adopt the Paczyński & Wiita potential to mimic its geometry.

The most difficult part of solving the global solution is the radial momentum equation 2. The key idea of simplifying the global ADAF solution is to replace this differential equation by the following simple algebraic relation:

$$\Omega = f\,\Omega_k \tag{16}$$

With,

$f = f_0$ for $r > r_{ms}$,

$$\frac{f_0\,\Omega_k r_{ms} {r_{ms}}^2}{r^2 \Omega_k}\left(\frac{r}{r_{ms}}\right)^n = f_0\,\frac{r-r_g}{2r_g}\left(\frac{r}{r_{ms}}\right)^{n-3/2} \text{ for } r < r_{ms}, \tag{17}$$

Where $r_{ms} = 3r_g \equiv GM / c^2$ is the innermost stable circular orbit.

The above simplification is based on the following physical consideration. The immediate idea to think of to simplify the radial momentum equation is to use the self-similar solution obtained. Consider

equations (7)-(9) in NY94, using equation (9) in NY94 to solve for ε in terms of the sound speed c_s:

$$5+2\varepsilon=2\frac{v_k^2}{c_s^2}$$

$$\frac{\Omega}{\Omega_k}=\left(1-\frac{5c_s^2}{2v_k^2}\right)^{1/2}=\frac{10-6\gamma}{9\gamma-5}=\text{constant} \quad (18)$$

Where γ is the adiabatic index.

Then substitute this in equations (8) and (7) in NY94 to obtain for a fully advection-dominated flow. Therefore, setting *f* in equation (16) to this constant. However, that the simplified solution is very sensitive to the value of *f*. The reason is that, as shown below, the radial velocity is sensitive to the value of *f*. The velocity determines the density and also the temperature via the energy equations, two quantities that determine the emitted spectrum. Therefore, set *f* as a free parameter that adjust for different accretion parameters $\dot{M}$ and α to get the best approximation.

Because the angular momentum in a global solution keeps decreasing with decreasing radius, while the Keplerian angular momentum begins to increase when, $r>r_{ms}$ *f* cannot be a constant when $r<r_{ms}$. Instead, we require the angular momentum to be continuous at r_{ms} and assume that it is proportional to $\left(\frac{r}{r_{ms}}\right)^n$

The quantity *j* is the specific angular momentum of the accretion gas when it falls into the black hole, and it is the eigenvalue of the exact global solution. In simplified model, setting *j* as the second free parameter and adjust its value to get the best approximate solution for *v*.

Substituting equations (1), (6), and (9) into the energy equations for ions and electrons, equations (4) and (5), we have two differential equations with two unknown variables, T_i and T_e. All other quantities such as v, ρ, c_s, and *H* can be expressed as simple functions of T_i and T_e for a given $\dot{M}$ and α and assumed values of the free parameters f_0 and *j*. When, T_i and, T_e are given at the outer boundary, one can easily integrate the differential equations inwards to get the approximate global solution.

Adjusting the values of f_0 and *j* for different $\dot{M}$ and α to get the best simplified global solution. That is all quantities such as ρ, v,T_e, T_i and the emitted spectrum, are very close to the exact global solution.

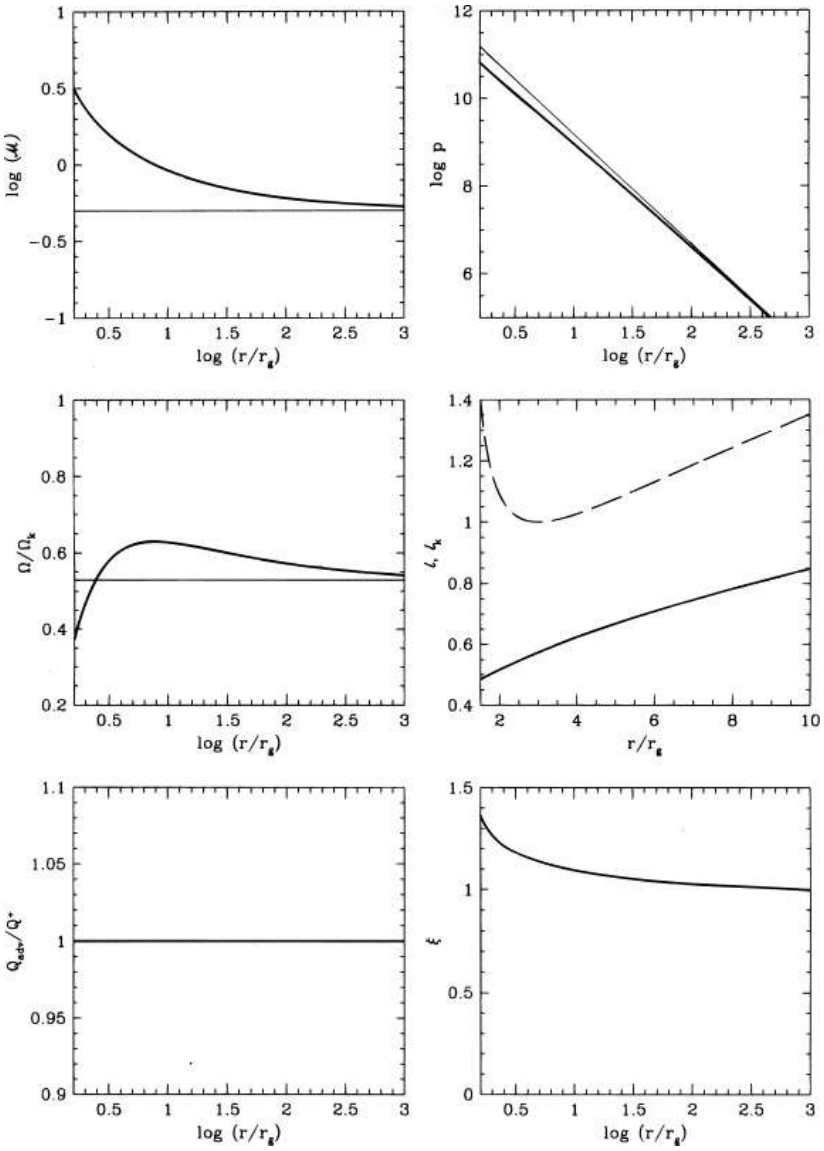


Figure 4: The heavy lines are the global solution and the thin lines are the corresponding self-similar solution

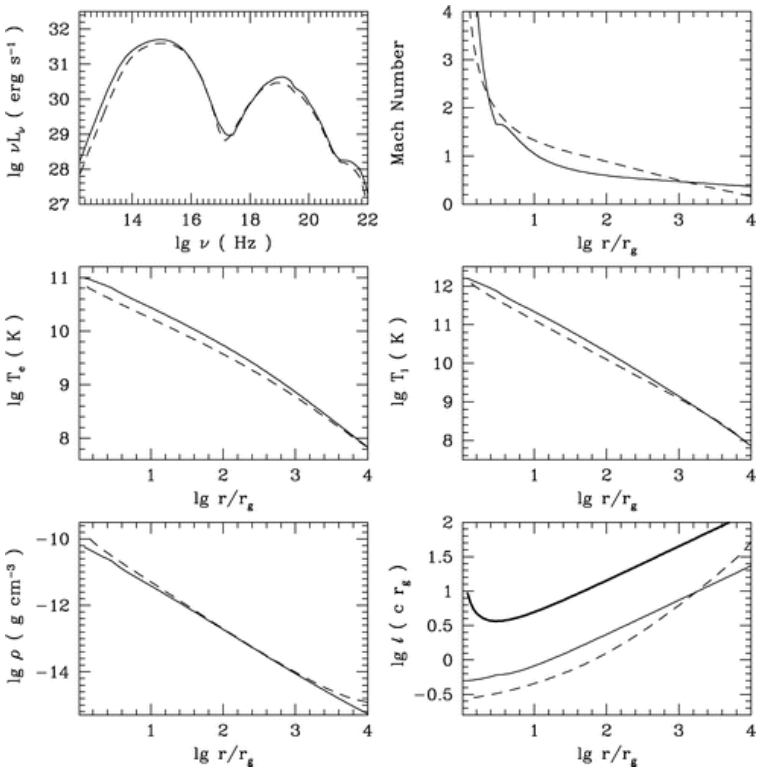


Figure 5: Diagram to differentiate Simplified (*solid line*) and exact (*dashed line*) global ADAF solutions

Properties of ADAF

- ADAFs are particularly good at generating powerful winds and relativistic jets.
- Large pressure and very high T_i, T_e (virial, since ADAF loses very little heat) geometrically thick: H/R ~c_s / v_k ~ 0.6 Sub-Keplerian rotation: $\Omega < \Omega_k$ Very low density The ADAF is thermally stable.
- ADAFs are Everywhere Thin disk systems are bright (high $\dot{M}$, high efficiency) and tend to dominate observational programs ADAFs are much fainter, and harder to observe, but they occupy a very large range of parameter space Probably >90% of BHs in the universe are in the ADAF phase.
- ADAFs around Stellar-Mass BHs ADAFs are found in Quiescent State Hard State.

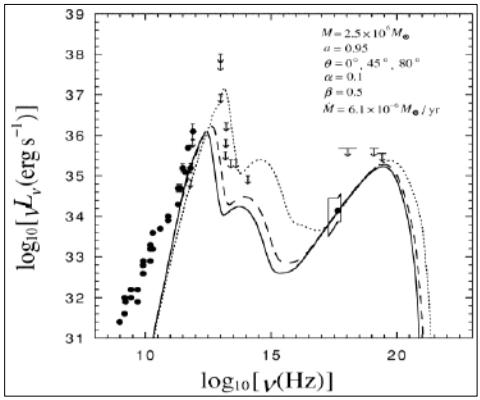
- This is found in Intermediate States But not the Steep Power Law State.
- Extreme inefficiency of Sgr A* and other quiescent BHs is explained.
- It can predict the existence of event horizon.

There are two kinds of ADAF

- Radiation trapped in the gas and cannot diffused out before gas falls into BH. "Slim disk" solution (Abramowicz, Czerny, Lasota & Szuskiewicz 1988) where L$\sim L_{EDD}$
- Gas is very dilute and can't radiate its thermal energy before its falls into the black hole. Radiatively inefficient - "RIAF" (ICHIMARU 1977, NY 1994, 95; Abramowicz *et al*. 1995) L$\sim(0.01-0.1)L_{EDD}$

Spectrum

The most fascinating feature of the ADAF model is its two-temperature structure, with the ion temperature being much higher than the electron temperature, which nicely explains the observed spectra of the black hole systems. The spectrum from an ADAF around a black hole ranges from radio frequencies 10^9Hz to gamma{ray frequencies > 10^{23}Hz, and can be divided into two parts based on the emitting particles: The radio to hard X ray radiation is produced by electrons via synchrotron, bremsstrahlung and inverse Compton processes (Mahadevan 1997). The gamma ray radiation results from the decay of neutral pi-ons created in proton-proton collisions (Mahadevan, Narayan & Krolik 1997).



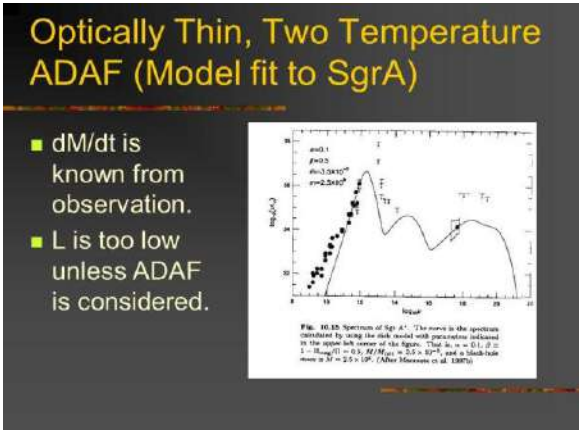


Figure 6: Spectrum analysis of Sgr A* using ADAF model

Conclusion

Advection-dominated accretion flow (ADAF) is an important type of solution for black hole accretion. A prominent feature of an ADAF compared to the standard thin disk is its low radiative efficient at low accretion rates. The self-similar solution may provide an excellent description of an advection-dominated flow far from the flow boundaries as it was shown by Narayan & Yi (1994).This self -similar solution can never be an asymptotic solution for an ADAF. However, the structure of the flow in the inner and outer regions may have an important influence on the general properties of the modelled objects. When we model black hole accretion sources as advection-dominated accretion flows (ADAFs), it is necessary to use the global solution to the equations rather than the simpler self-similar solution, since the latter is inaccurate in the region near the black hole where most of the radiation is emitted. Yet there are some problems with this model like while ADAFs nicely explain the absence of emission, it is not clear whether their spectra are actually observed. However, it is technically very difficult to calculate the global solution of an ADAF. An ADAF is transonic, and thus its global solution should satisfy the sonic-point condition in addition to the outer boundary condition. Mathematically, it is a two-point boundary value problem and not easy to deal with. This is an obstacle to the wide application of the ADAF model.

The Spectrum of Sgr A* (radio and X-ray emission) is well explained by a jet model while ADAFs neither reproduce the radio nor the X-ray spectrum well. Even if jet models are sometimes more successful in explain observed

spectra, still we need an ADAF to explain the absence of disk emission and some other properties. Another attractive feature of an ADAF is that it can partly solve the problem of the origin of X-ray emission from accretion flows.

References

1. Shakura N.I., Sunyaev R.A., 1973, A&A, 24, 337
2. Narayan R., Yi I., 1995, Advection-dominated accretion: underfed black holes & neutron stars, astrophysics. J., 452, 710
3. Advection-dominated accretion: Global transonic solutions, Chen, Abramowicz and Lasota
4. Accretion disc theory: G.S. Bisnovatyi-Kogan
5. A simplified global solution for Advection-dominated accretion flow: Yuan, Feng; Ma, Renyi; Narayan, Ramesh, 2008
6. Self-similar solution for Advection-dominated accretion flow: Ramesh Narayan1994
7. https://en.wikipedia.org/wiki/Accretion_disk
8. https://iopscience.iop.org/article/10.1086/308768/fulltext/40355.text.html
9. Physics of AGN: Dr Heino Flack

Chapter - 12

Retailer's Inventory Decisions for Deteriorating Items with Exchange Scheme Under the Effect of Advertisement Efforts with Shortages

Authors

Ayan Chakraborty
Research Scholar, Department of Mathematics, Techno India University, West Bengal, India

Nirmal Kumar Duari
Assistant Professor, Department of Mathematics, Techno India University, West Bengal, India

Tripti Chakrabarti
Prof. & Dean of Basic Science, Department of Mathematics, Techno India University, West Bengal, India

Subhabrata Mondal
Department of Mathematics, Swami Vivekananda University, Barrackpore, West Bengal, India

Santu Ghorai
Assistant Professor, Department of Mathematics, Swami Vivekananda University, Barrackpore, West Bengal, India

Chapter - 12

Retailer's Inventory Decisions for Deteriorating Items with Exchange Scheme Under the Effect of Advertisement Efforts with Shortages

Ayan Chakraborty, Nirmal Kumar Duari, Tripti Chakrabarti, Subhabrata Mondal and Santu Ghorai

Abstract

This study looked at an inventory model with exchange scheme policy and product deterioration. Customers may purchase used goods under an exchange program in exchange for a discount on new purchases. Products' ongoing degradation is taken into account. Retailers aim to use advertising to bring in as many customers as possible. Customers may also purchase new products at regular prices. Retailers invest money on inspection to assess the quality of used goods. If a product can be reused, then it can be sold for less than what it originally cost; otherwise, it is sold for nothing. Shortages are taken into account in this model. The study's goal is to reduce the retailer's overall costs. Numerical analysis and sensitivity analysis are used to justify the model.

Keywords: Exchange scheme, inspection, rework, garbage value, deterioration, shortage cost.

1. Introduction

Storage and maintenance of product demand, procurement decisions, competitive market strategies, customer happiness, promotional rules, and maintenance of product quality are just a few of the many components that make up inventory management, business standing, financial expansion, etc. Inventory managers must make earnest efforts to withstand and, as a result, expand the firm in a competitive market. The adoption of various promotional initiatives has been the subject of ongoing analysis by market players. Promotional strategies aim to boost customer awareness of the product at all levels, as well as customer traffic and deal conversion. Sellers

of many different products, including televisions, refrigerators, air conditioners, bikes, vehicles, books, gold jewellery, mobile phones, computers, and laptops, have recently expressed a lot of interest in and implemented its exchange program. Many customers are reluctant to part with their old possessions. The exchange program's policy of returning the old and taking home the new encourages customers to purchase new goods. The goal of promotional methods is to increase deal conversion, customer traffic, and product awareness among all customer segments. Recently, sellers of a wide range of items, including televisions, refrigerators, air conditioners, bikes, bicycles, vehicles, books, gold jewellery, mobile phones, computers, and laptops, have shown a lot of interest in and accepted its exchange scheme. Many clients find it difficult to part with their old belongings. Customers are encouraged to buy new products because of the exchange program's rule requiring them to return the old and take home the new. Customers who surrender outdated products have the chance to receive a tempting discount on the purchase of new goods rather than having to sell them for scrap. The clients so benefit from the arrangement. The vendor, on the other hand, benefits from the scheme's increased sales volume. It draws more football fans to the store. As part of the program, pool objects are also assessed before being separated into those in good shape and those in poor condition. Good-quality products are fixed and made available as reconditioned goods. For a little fee, manufacturers will accept used, defective products back for research and development purposes or for the parts to be used in other, more useful gadgets that require repair. In order to make money, it is also possible to dispose of old, damaged goods at their rubbish value. The exchange program benefits both the customer and the store, which might increase their profits as a result of the program. As a result, both the retailer and the customer come out ahead in this case. It's important to spread knowledge about the program to the general public in order to benefit as much as possible from such program.

2. Literature review

By effectively promoting a product, consumers may be made eloquently aware of promotional tactics. (1) Investigated the effects of investment in preservation technology on a joint location inventory problem for non-immediately deteriorating items. They investigated a promotional technique that combined repeated incremental reductions with a tolerable payment delay to achieve the best outcomes in the inventory system. The Ghare & Schrader, Covert & Philip, Dave & Patel, and 1981 Dave & Patel model with

constant and variable degradation rate and quantity discount was explored and developed in the year 1963. Nowadays, a lot of focus is placed on how the basic idea of preservation technology has drastically decreased the rate of deterioration. This essay explores the idea of preservation technology because it can lessen the rate of deterioration in a time of significant societal change. Additionally, for products that degrade quickly, the rate of deterioration is extremely sensitive to sales, inventory, and order numbers. Advanced rates of deterioration could result in greater yearly relevant costs and a slower rate of demand (Yang & Wee, 2006; Johnny *et al*., 2007). According to Ouyang *et al*. (2006), the total annual relevant inventory cost can be greatly decreased if the retailer is able to slow down the rate at which the goods deteriorate by building the storage facility.

Numerous businesses have made significant investments in machinery for this purpose, extending the product's shelf life. For instance, the rate of fruit, flowers, and seafood deterioration in supermarkets has been reduced by the use of chilling equipment. Morris and Murr's 1975 study found that low temperatures can prolong storage life and inhibit decomposition. Wee *et al*. (2010) used preservation techniques to develop a model.

Huang *et al*. addressed the concepts about preservation technology investment for degrading inventory (2011). Dye and Hsieh (2012) developed an effective replenishment method for failing items via PT investment. The impact of PT investment on a model of non-instantaneous inventory deterioration was studied by researchers Duari *et al*. (2019), Duari *et al*. (2022) and Dye (2013). Hsieh and Dye (2013) developed a production inventory model that takes into account the outcomes of PT investments when demand varies over time and Duari *et al* (2012) has adopted advertisement with shortages.

In a supply chain system, coordination between the manufacturer and distributor is an effective way to meet demand. Supply chain management was seen to use an integrated strategy. This is how the EOQ model of commodities with deteriorating Weibull distribution, shortages, and trended demand is viewed by T. Chakrabarti, B.C. Giri, and K.S. Chaudhury (1998). Chen and Kang developed the linked vendor-buyer inventory models with various permissible payment delays in 2007. Rau and Ouyang investigated an integrated production inventory policy of the ideal batch size in a supply chain (2008). S. Sarkar and T. Chakrabarti developed a backlog-allowing inventory model for immediately deteriorating products in 2012. Urvashi *et*

al. developed a supply chain model for items that deteriorate under inflation while using preservation techniques (2014). Urvashi *et al*. developed the supply chain model in three distinct contexts: 2008, 2009, and 2013. Nita H. Shah *et al*. (2021) developed a model of Retailer's inventory decisions with exchange scheme under the effect of advertisement efforts.

3. Assumptions and notations

Notations

Y: Product demand rate under no exchange category.

Y_{ex}: Product demand rate under exchange scheme.

a: Constant scale demand for demand function with no exchange category($a>0$).

a_1: Constant scale demand for demand function with exchange category($a_1>0$).

b: Price elasticity factor for demand function with no exchange category ($b>0$)

b_1: Price elasticity factor for demand function with exchange category ($b_1>0$)

α: Advertisement scale factor ($\alpha>1$)

β: Advertisement effectiveness factor($\beta>0$)

Costs related to the model:

Z: Set up cost or ordering cost per order.

c: Purchase cost per unit.

h: Holding cost per unit per unit time

i: Inspection cost per unit for old products.

r: Rework cost per unit

HC: Total holding cost of inventory.

IC: Total inspection cost of the inventory.

MC: Total advertisement cost of the inventory.

RC: Total rework cost of good condition old inventory.

TC: Total cost of inventory for the retailer.

d: Selling price reduction factor under exchange scheme.

g: Garbage value per unit for useless old products.

$I_w(t)$: Inventory level of product under without exchange scheme at time t.

$I_{ex}(t)$: Inventory level of product under with exchange scheme at time t.

$I_{old}(t)$: Inventory level of old products received under exchange scheme at time t.

P_e: Reduced selling price per unit under exchange scheme.

Assumptions

1) This model has been developed in the interval cycle of (0, X). Advertisement Cost is considered to increase the sales of the product and demand of the product.
2) Without an exchange system, product demand is calculated as $Y=(a-bp)\ (\alpha+\beta M)$. $Y_{ex}=(a_1-b\ _1p)\ (\alpha+\beta M)$ is used to represent product demand in exchange schemes.
3) Old products are sent for inspection process.
4) In this paper one modification has been done that shortages are taken into consideration.

4. Mathematical inventory model

Customers have the option of purchasing new products with or without an exchange plan. Therefore, this model can be discussed in two parts. The following can be written as the retailer's inventory decisions under an exchange program or without one, subject to the influence of advertising, with ongoing deterioration and shortages.

Case 1: Without exchange scheme

In this scheme Product Demand rate $R=(a-bp)(\alpha+\beta M)$

$$\frac{d(Iw(t))}{dt} = -Y - \theta \quad 0\leq t\leq t_1 \qquad (1)$$

$$\frac{d(Iw(t))}{dt} = -Y \quad t_1\leq t\leq X \qquad (2)$$

Case 2: With exchange scheme

In this scheme Product Demand rate $R_{ex}=(a_1-b_1p_e)(\alpha+\beta M)$

$\frac{d(Iex(t))}{dt}$ = -Y(X-t) - θ $0 \le t \le t_1$ (3)

$\frac{d(Iex(t))}{dt}$ = -Y(X-t) $t_1 \le t \le X$ (4)

Now from case 1, equation no (1) using the initial conditions $I_w(t_1)=0$, $I_w(t)=Q_{w1}$ we get

$Q_{w1}= (Y+\theta)(t_1-t)$ $0 \le t \le t_1$ (5)

Case 1 equation no (2) using the initial condition $I_w(t_1)=0=Q_{w2}$ we get $t^2-t_1^2$

$Q_{w2}= Y(t_1-t)$ $t_1 \le t \le X$ (6)

Now from case 2, equation no (3) using the initial conditions $I_{ex}(t_1)=0$ and $I_{ex}(t)=Q_{w3}$

$Q_{w3} = Y_{ex}(t^2-t_1^2)/2 +\theta(t_1-t)$ $0 \le t \le t_1$ (7)

Now from case 2, equation no (4) using the initial conditions $I_{ex}(t_1)=0$ and $I_{ex}(t)=Q_{w4}$

$Q_{w4} = Y_{ex}(t^2-t_1^2)/2$ (8)

Total ordering quantity in the interval $0 \le t \le t_1$ is

$Q_1= Q_{w1}+ Q_{w3} = (Y+\theta)(t_1-t) + Y_{ex}(t^2-t_1^2)/2 +\theta(t_1-t)$ (9)

Total ordering quantity in the interval $t_1 \le t \le X$ is

$Q_2= Q_{w2}+ Q_{w4} = Y(t_1-t)+ Y_{ex}(t^2-t_1^2)/2$ (10)

Additionally, the shopkeepers receive an old product for each new product they sell under the exchange program. As a result, the differential equation governs the system's inventory levels of outdated goods at any given time,

$\frac{d(Iold(t))}{dt}$ = Y_{ex}, $0 \le t \le X$ (11)

Using initial condition $Iold(0)=0$ to solve equation (11) the inventory level of old products can be represented as $Iold(t)= Y_{ex}t$ (12)

The quantity of old products received in the system is $Q_{old} = R_{ex}X$ (13)

Ordering cost of the retailer (OC)= Z (14)

Total Purchase Cost (PC)= $CQ=C(Q_{w1}+ Q_{w3})=C((Y+\theta)(t_1-t) + Y_{ex}(t^2-t_1^2)/2 +\theta(t_1-t))$ (15)

Holding Cost(HC)= $h\int_0^{t1} Q1\,dt = h[(Y+\theta)t_1^2/2 - Y_{ex}t_1^3/3 +\theta\, t_1^2/2]$ (16)

Deterioration Cost(DC)= $\theta\,[(Y+\theta)t_1^2/2 - Y_{ex}t_1^3/3 +\theta\, t_1^2/2]$ (17)

Shortage Cost (SC)= $S[Y(t_1X-X^2/2)-Yt_1^2/2 + Y_{ex}(X^3/6 -Xt_1^2/2)+Y_{ex}\, t_1^3/3]$ (18)

Advertisement Cost(AC)= $M[(a-bp)(\alpha+\beta M)+(a_1-b_1p_e)(\alpha+\beta M)]t_1$ (19)

Inspection Cost(IC)= $i\,(a_1-b_1p_e)(\alpha+\beta M)X$ (20)

Rework Cost(RC) = $r\,\square(a_1-b_1p_e)(\alpha+\beta M)X$ (21)

From above equation we can get total cost incurred to the retailer is

TC = OC+PC+HC+DC+SC+AC+IC+RC

= $Z+ C((Y+\theta)(t_1-t) + Y_{ex}(t^2-t_1^2)/2 +\theta(t_1-t))+ h[(Y+\theta)t_1^2/2 - Y_{ex}t_1^3/3 +\theta\, t_1^2/2]$

$+ \theta\,[(Y+\theta)t_1^2/2 - Y_{ex}t_1^3/3 +\theta\, t_1^2/2]+ S[Y(t_1X-X^2/2)-Yt_1^2/2$

$+ Y_{ex}(X^3/6 -Xt_1^2/2)+Y_{ex}\, t_1^3/3]+ M[(a-bp)(\alpha+\beta M)+(a_1-b_1p_e)(\alpha+\beta M)]t_1$

$+ i\,(a_1-b_1p_e)(\alpha+\beta M)X+ r\,\square(a_1-b_1p_e)(\alpha+\beta M)X$

5. Numerical and sensitivity analysis

For numerical and sensitivity analysis we have taken following values of the variables which are under mentioned as: Z=300; a=700; b=15; a1=400; b1=7; c=12; h=0.5; i=.2; p=0.5; r=2; Alpha=1; Beta=0.06; Gamma=0.05; Y=5; Theta=0.5; t1=5; t=2; X=8; REX=12; S=5; pe=1;M=5;

We have used lingo software for the numerical results. We got global optimum solution of this problem. The Global optimum value of the cost function is TC= 37985.54.

We took different values of the variables and increase and decrease the values of the variables as 50% and 20% respectively. We have seen that total cost changed as we changed the values of the variables. We are interested to find out the variables which are sensitive for this model. From the sensitivity analysis we can say that few variables are highly sensitive like advertisement cost and Inspection Cost etc. Sensitive Analysis of different variables are analyzed with the below mentioned graphs.

Variables	Changed values of the variables	Changed values of the cost function	% changed in cost function
Demand Without Exchange Scheme(Y)	15(Hypothetical)	38470.5	.63% increase
	12(Hypothetical)	38325	.25% increase
	10(36)	38228	Model value
	8(Hypothetical)	38131	.25% decrease
	5(Hypothetical)	37985.5	.63% increase
Shortage Cost(S)	7.5(Hypothetical)	38418	.49% increase
	6(Hypothetical)	38304	.19% increase
	5(36)	38228	Model Value
	4(Hypothetical)	38152	.19% decrease
	2.5(Hypothetical)	38038	.49% decrease
Advertisement Cost(M)	7.5(Hypothetical)	62114.8	62.48% increase
	6(Hypothetical)	47294.3	.23.71% increase
	5(36)	38228	Model Value
	4(Hypothetical)	29813	22.01% decrease
	2.5(Hypothetical)	18411.9	51.83% decrease
Inspection Cost(i)	.3(Hypothetical)	38636.8	1.06% increase
	.24(Hypothetical)	38391.5	.42% increase
	.2(36)	38228	Model Value
	.16(Hypothetical)	38064.5	.42% decrease
	.1(Hypothetical)	37819.3	1.06% decrease

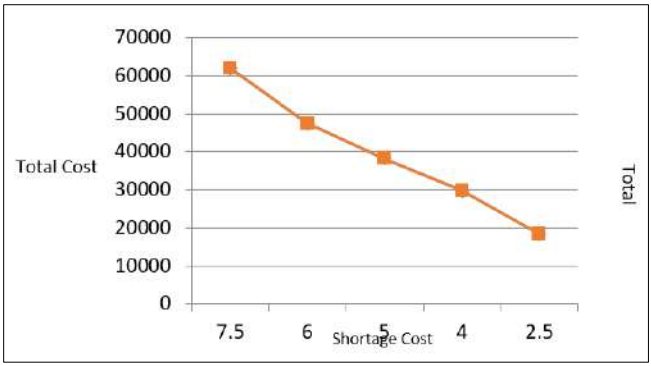


Figure 1: (Hypothetical) represents the relationship between shortage cost and total cost

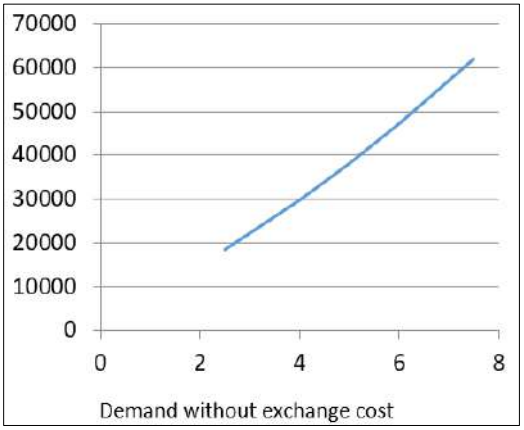


Figure 2: (Hypothetical) represents the relationship between Demand without exchange and total cost

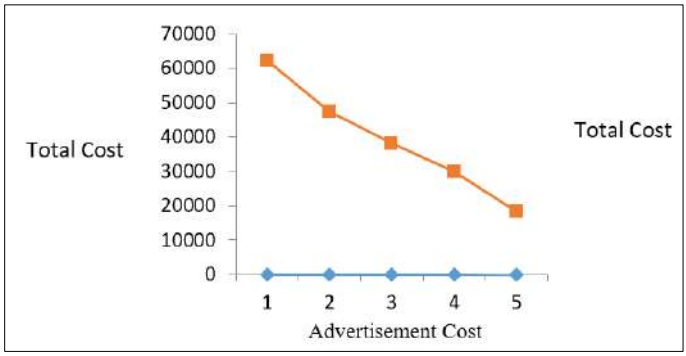


Figure 3: (Hypothetical) represents the relationship between advertisement cost and total cost

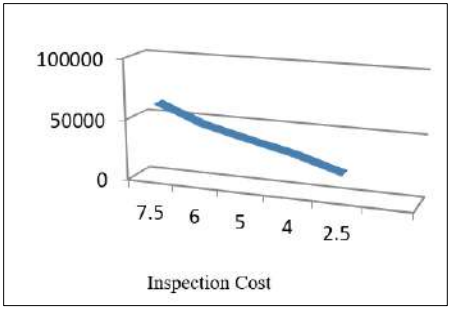


Figure 4: (Hypothetical) represents the relationship between inspection cost and total cost

From the above graph we can say that the variables are very much sensitive to the cost function.

Data availability

I took the data from Nita H Shah's Paper which I cited in [35]. Rest of the data are taken after increasing or decreasing the value 50% and 20% respectively. These data are called as hypothetical.

6. Conclusion

This paper focuses on cost minimization of the retailer. Retailer gives offers to customer to buy new product with reduce price advantage and exchange scheme. Retailer uses advertisement for increase in demand of the product. As a consequences retailer has to pay advertisement cost but it has a great impact in increasing in demand. Shortages are considered in this model. An old product which has good condition has been sold by the retailer after reworking and molding the product. For inspection retailer has to pay inspection cost. This paper is very much realistic. But this paper can be modified after taking consideration the idea of fuzzy and preservation technology.

References

1. Bhunia, A.K., Kundu, S., Sannigrahi, T. and Goyal, S.K. (2009). An application of tournament genetic algorithm in a marketing oriented economic production lot-size model for deteriorating items. *International Journal of Production Economics*, 119(1) 112-121.
2. Bansal, Kapil. (2013). Effect of inflation on supply chain model with exponential demand. Multi Disciplinary Edu Global Quest, January 2013
3. Chang, H.J. and Lin, W.F. (2010). A partial backlogging inventory model for non-instantaneous deteriorating items with stock-dependent consumption rate under inflation. *Yugoslav Journal of Operation Research*, 20(1) 35-54.
4. Chang, C.T., Teng, J.T. and Goyal, S.K. (2010). Optimal replenishment policies for non- instantaneous deteriorating items with stock-dependent demand. *International Journal of Production Economics*, 123(1) 62-68.
5. Chen L.H, Kang F.S, (2007) 'Integrated vendor buyer cooperative inventory models with variant permissible delay in payments', European Journal of Operational Research, 218(1), 106-112.

6. Covert, R.P. and Philip, G.C. (1973). An EOQ model for items with Weibull distribution deterioration. *AIIE Transaction*, 5(4) 323-326.

7. Ghare, P.M. and Schrader, G.F. (1963). A model for exponential decaying inventory. *Journal of Industrial Engineering*, 14(6) 238-43.

8. Goyal, S.K. and Giri, B.C. (2001). Recent Trends in Modeling of Deteriorating Inventory. *European Journal of Operational Research*, 134(1) 1-16.

9. Dave, U. and Patel, L.K. (1981). (T, Si) policy inventory model for deteriorating items with time proportional demand. *Journal of the Operational Research Society*, 32(1) 137-142.

10. Dye, C.Y. (2013). The effect of Preservation Technology investment on a non- instantaneous deteriorating inventory model. *Omega*, 41(1) 872-880.

11. Duari *et al*. (2012) "A Marketing Decision problem in a periodic review stochastic model with shortages", *International Organization of Scientific Research* Vol. 1 (6) Pp. 35-38, Jul-Aug 2012

12. Duari *et al*. (2019) "Non-instantaneous Deteriorating Model for Stock-Dependent Demand with Time-Varying Holding Cost and Random Decay Start Time", Published in book: Optimization and Inventory Management, In August 2019, DOI: 10.1007/978-981-13-9698-4_9, Springer

13. Duari *et al*. (2022) A Supply Chain model for deteriorating items Under Inflation with Preservation Technology and Shortages, Gradiva review journal, volume 8 issue 7 2022, page no: 1371, issn no: 0363-8057

14. Dutta, D. and Kumar, P. (2013). Fuzzy inventory models for deteriorating items with shortages and fully backlogged condition. *International Journal of Soft Computing and Engineering*, 3(2), 393-398.

15. Huang, Y.H., Wang, C.C., Huang, C.J. and Dye, C.Y. (2011). Comments on preservation technology investment for deteriorating inventory. *African Journal of Business Management.* 5(11) 4636-4643.

16. Hsieh, T.P. and Dye, C.Y. (2013). A production inventory model incorporating the effect of preservation technology investment when demand is fluctuating with time. *Journal of Computational and Applied Mathematics*, 239, 25-36.

17. Johnny, C.H., Adriano, O.S. and Chang, Y.L. (2007). An evaluation of lot-sizing heuristics for deteriorating inventory in material requirements planning systems. *Computers and Operations Research*, 34(9) 2562-2575.

18. Kang, S. and Kim, I. (1983). A study on the price and production level of the deteriorating inventory system. *International Journal of Production Research*, 21(6) 449-460.

19. Murr, D.P. and Morris, L.L. (1975). Effect of storage temperature on post change in mushrooms. *Journal of the American Society for Horticultural Science*, 100(1) 16-19.

20. Ouyang, L.Y., Wu, K.S. and Yang, C.T. (2006). A study on an inventory model for non- instantaneous deteriorating items with permissible delay in payments. *Computers and Industrial Engineering*, 51(4) 637-651.

21. Ruoning, X. and Xiaoyan, Z. (2010). Analysis of supply chain coordination under fuzzy demand in a two-stage supply chain. *Applied Mathematical Modeling*, 34(1) 129-139

22. Singh, S.R., Kumari, R. and Kumar, N. (2011). Optimization of fuzzy inventory model for differential items. *International Journal of Operational Research*, 11(3) 290-315.

23. Singh, C. and Singh, S.R. (2011). Imperfect production process with exponential demand rate, Weibull deterioration under inflation. *International Journal of Operational Research*, 12(4) 430-445.

24. Rau H., Ouyang B.C.(2008)' An optimal batch size for integrated production inventory policy in a supply chain', European Journal of Operational Research, Vol.185,No.2.,pp.619-634

25. S. sarkar and T. Chakrabarti: (2012); "Inventory Problem of Instantaneous Deteriorating Items with fuzzy Ramp-type Demand allowing Backlogging". Bulletin of the Calcutta Mathematical society, Volume 104, No. 3, (pp: 219-236).

26. Singh,S.R.,Urvashi,(2008),'An integrated production inventory model with Expiration date and Uncertain Lead Time in a Supply Chain', International Transactions in Mathematical Sciences and Computer Application, VOL-1,No-1,pg 187-202.

27. Singh,S.R.,Urvashi,(2009), 'Supply Chain Model with Imperfect Production Process and Stochastic Demand under Chance and Imprecise Constraints with Variable Holding Cost'. The ICFAI University Journal of Computational Mathematics, Vol-2, No-1,pg-37-62.

28. Singh,S.R.,Urvashi,(2013),'Green supply chain model with product remanufacturing under volume flexible environment', Procedia Technolgy 10(2013),pg216-226,Elsevier Itd.

29. S.R. Singh and Vandana Gupta: Supply chain production model with preservation technology under fuzzy environment. June 2014, International Journal of Industrial Engineering Computations 5(3):459-472. DOI:10.5267/j.ijiec.2014.3.002

30. T. Chakrabarti, B.C. Giri, K. S. Chaudhury: (1998), "An EOQ model of items with deterioration of Weibull distribution, shortages and trended demand": An Extension of Philip's model. Computers and Operations Research, U.S.A. (International journal).

31. Urvashi, S.R.Singh, Vikas Sharma; (2014) "A Supply Chain Model for the Items with deterioration under Inflation with Preservation Technology"ICRTET2014.

32. Wee, H.M., Teng, H.M. and Hsu, P.H. (2010). Preservation technology investment for deteriorating inventory. *International Journal of Production Economics*, 124(2) 388-394.

33. Yang, P.C. and Wee, H.M. (2006). A collaborative inventory system with permissible delay in payment for deteriorating items. *Mathematical and Computer Modeling*, 43(3-4) 209-221.

34. Bhula B, Dash JK, Rajendra Kumar M. An optimal inventory model for permissible items under successive price discounts with permissible delay in payments. Opsearch.2019;56(1):261-281.

35. Nita H. Shah, Prantik H. Shah, Milan B Patel. Retailer's Inventory Decisions with Exchange Scheme under the effect of Advertisement efforts, Taylor & Francis.2021.

36. Kundu.S & Chakrabarti.T [2018] Joint optimal decisions on pricing and local advertising policy of a socially responsible dual-channel supply chain.American Journal of Mathematical anf Management Sciences, 37(2),117-143.

37. Kundu.S & Chakrabarti. T [2018] Impact of carbon emission policies on manufacturing, remanufacturing and collection of used item decisions with price dependent return,OPSEARCH,55(2),532-555.

38. Kundu.S & Chakrabarti.T [2019] A Fuzzy rough integrated multi stage supply chain inventory model with carbon emissions under inflation and time value of money. International Journal of Mathematics in operation research, Inderscience,14(1),123-145.

39. S Kundu and T Chakrabarti: (2018); "A fuzzy rough integrated multi-stage supply chain inventory model with carbon emissions under inflation and time-value of money", International Journal of Mathematics in Operational Research, In- derscience, Vol 14 Issue 1.

Chapter - 13

Impact of Carbon Emission on EPQ Model in the Context of Rhino Bricks Production

Authors

Ayan Chakraborty
Research Scholar, Department of Mathematics, Techno India University, West Bengal, India

Nirmal Kumar Duari
Assistant Professor, Department of Mathematics, Techno India University, West Bengal, India

Tripti Chakrabarti
Prof. & Dean of Basic Science, Department of Mathematics, Techno India University, West Bengal, India

Subhabrata Mondal
Department of Mathematics, Swami Vivekananda University, Barrackpore, West Bengal, India

Santu Ghorai
Assistant Professor, Department of Mathematics, Swami Vivekananda University, Barrackpore, West Bengal, India

Chapter - 13

Impact of Carbon Emission on EPQ Model in the Context of Rhino Bricks Production

Ayan Chakraborty, Nirmal Kumar Duari, Tripti Chakrabarti, Subhabrata Mondal and Santu Ghorai

Abstract

Carbon dioxide emission is the main reason behind the global warming and Plastic disposition encourages to increase carbon dioxide emission. Environmental decarbonization is very challenging issue now a days. Reduction of uses of plastic is the way to sustain the green environment. Sources of plastic are from the plastic from the industry and plastic which is being gathered by the consumers. Plastic used by the consumer are accumulated in the dump yard. Sand became toxic because of sand casting repeatedly. Plastic that are collected both from the industry and consumer are taken and they are brought to the rhino brick industry. Toxic sand and plastic are mixed in a finite temperature so that it becomes the shape of rhino bricks.

Keywords: Decarbonization, deterioration, carbon emission, sustainability.

1. Introduction

Emission of greenhouse gases from various industries, production, wetlands, and inventories are currently the biggest obstacles for policymakers trying to sustain our developments. One of the main causes of our lovely planet's warming temperatures is carbon emissions from the disposal of plastic. Reusing plastic is one of the key initiatives that are made to reduce carbon emissions. Reusing plastic is a great way to lower carbon emissions in public areas and is the most effective way to slow global warming. Typically, plastic that was stored from various sources is burned outside, releasing noxious fumes into the atmosphere. But if we use that plastic in the rhino brick industry that pollution can be reduced. Another thing different metal shapes are prepared after putting melted high temperature metal into sand and that process is called sand casting. After repeated production of different metal sand became toxic. Toxic sand

increase the toxicity of the wet lands and reduces the fertility of the land. This sand can be reuse in the rhino brick industry. In rhino brick industry plastic materials that are collected either from plastic industry or from household are cutting into small pieces. Then sand and small pieces of plastic are sent to a high temperature channel so that to form a mixture that are the process of rhino brick production which reduces the carbon emission in the environment.

Recently, green policies have been considered as a way to reduce carbon emissions, and Mashud *et al*.(2021a,b) established an inventory model. To reduce carbon emissions and greenhouse farms, Mashud *et al*. (2021a) suggested using sustainable inventory models. A green inventory model was created by Ritha *et al*. (2018) and Masud *et al*. (2020) to maintain the level of carbon emission. To lower the level of carbon emissions in the environment, Dutta (2017), Dutta, *et al*. (2019), and Dutta (2020) studied production and consumption inventory modelling. A model was created by Dieguez *et al*. (2021) to lower emissions in the energy sector.

N Bhattacharjee *et al*. (2022) [9] developed a model of decarbonization through production of rhino bricks from the waste plastic: EPQ model.

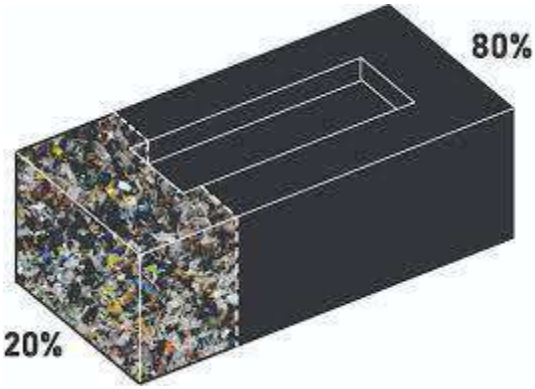


Figure 1: Represents the picture of rhino bricks from inside

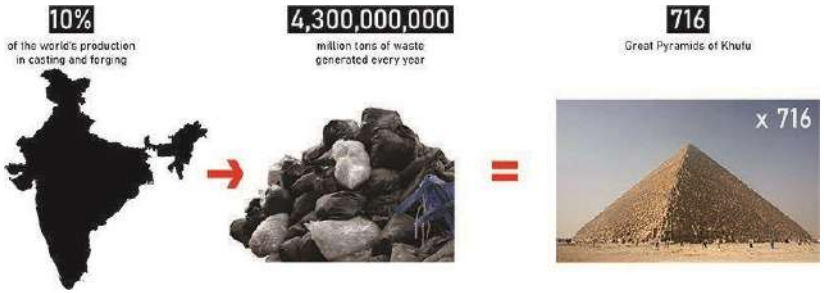


Figure 2: Represents the scenario of plastic pollution

2. Motivation of the work

In present situation global warming is one of the major threat to the earth and their inhabitants. Several factors that are responsible for global warming they are releases of CO_2, CFC, pollution particles, deforestation, Burning of different plastic related materials that releases unhygienic gases in our environment etc. We have already taken various steps to reduce green house gases but we are still working to stop the uses of plastic. Plastic is bio non degradable product which can be destroyed but we can change the recycling of plastic. One of such most impactful process to recycle plastic product is Rhino bricks production. Plastic that are collected from different sources that may be from societies or may be from plastic industry and they are cutting into small pieces and after sending those huge amount of plastic into a large process makes rhino bricks which reduces carbon emissions in the environment. Production of rhino bricks is also cost effective which has also satisfy the financial aspect.

3. Notations and assumptions

3.1 Notations

The following notations are used in the model.

i) I(t): Inventory level of plastic per unit per unit time.

ii) θ : Constant rate of deterioration, $0< \theta <1$

iii) ϕ : Rate of defective bricks.

iv) h: Holding cost per unit per unit time of plastic.

v) h_{rb}: Holding cost per unit per unit time of rhino bricks.

vi) c: Deterioration cost per unit per unit time of the plastic.

vii) c_{rb}: Deterioration per unit per unit time of the rhino bricks

viii) c_{e1}: Carbon tax in production per unit per unit time of the plastic.

ix) c_{e2}: Carbon tax in production per unit per unit time of the rhino bricks.

x) c_{eh}: Carbon tax in holding per unit per unit time of the rhino bricks.

xi) α: Fixed production cost (PC) of plastic.

xii) α_{rb} : Fixed production cost (PC) of rhino bricks.

xiii) v : Labour cost in the production of plastic

xiv) v_{eb} : Labour cost in the production of rhino bricks.

xv) s: Tool die cost in the production of plastic.

xvi) s_{rb} : Tool die cost in the production of rhino bricks.

xvii) c_t: Transportation cost per unit per unit time.

xviii) p: Salvage cost of defective bricks.

xix) T: production period of plastic.

xx) T_1: Production period of rhino bricks.

xxi) I_1: Initial inventory at t=0.

3.2 Assumption

The following assumption are taken into consideration.

i) There is no lead time.

ii) The factory's P1 plastic production rate and P2 rhino brick production rate are both constant. Although 50% of the P2 used in the manufacture of rhino brick comes from a landfill.

iii) Demand for rhino brick and plastic factories is constant.

4. Model development

Plastic factory

The inventory of plastic in the factory for the period $[0, T]$ is given by

$$\frac{dI}{dt} + \theta I = a(\gamma - 1)e^{bt} \cdot \tag{9.1}$$

Where $I(0)=I_1$

The inventory level in plastic factory during the time period $[0,T]$ is given by

$$I(t)= I_1 e^{-\theta t} + \frac{a(\gamma-1)}{I}(1+bt-\frac{b}{I}) \quad (9.2)$$

The cost associated with the supply chain to maintain the production factory

Setup Cost:

SC=O (9.3)

Holding cost:

$$HC= h\left[I_1(\frac{e^{-\theta T}}{-\theta}+\frac{1}{\theta}) + \frac{a(\gamma-1)}{I}(T+\frac{bT^2}{2}-\frac{bT}{I}) \right] \quad (9.4)$$

Deterioration Cost:

$$c\theta\int_0^T I(t)dt$$

$$= c\theta\int_0^T (I_1 e^{-\theta t} + \frac{a(\gamma-1)}{I}(1+bt-\frac{b}{I}))dt$$

$$= c\theta\left[I_1(\frac{e^{-\theta T}}{-\theta}+\frac{1}{\theta}) + \frac{a(\gamma-1)}{I}(T+\frac{bT^2}{2}-\frac{bT}{I}) \right]$$

Production Cost (PC)

$$(\alpha P_1 + v + sP_1^2)T$$

Carbon Tax in Production:

$$CTP = C_{e_1} P_1 T$$

Carbon Tax in Holding:

$$CTH= C_{eh} h\,(I_1(\frac{e^{-\theta T}}{-\theta}+\frac{1}{\theta}) + \frac{a(\gamma-1)}{I}(T+\frac{bT^2}{2}-\frac{bT}{I}))$$

Transportation Cost:

$$\text{TC}= C_t\left[I_1(\frac{e^{-\theta T}}{-\theta}+\frac{1}{\theta})+\frac{a(\gamma-1)}{I}(T+\frac{bT^2}{2}-\frac{bT}{I}) \right]$$

Overall Production Cost in the period $[0,T]$ is

OCPF(Overall Cost for Plastic Factory) = O + $h\left[I_1(\frac{e^{-\theta T}}{-\theta}+\frac{1}{\theta})+\frac{a(\gamma-1)}{I}(T+\frac{bT^2}{2}-\frac{bT}{I}) \right]$ + $c\theta\left[I_1(\frac{e^{-\theta T}}{-\theta}+\frac{1}{\theta})+\frac{a(\gamma-1)}{I}(T+\frac{bT^2}{2}-\frac{bT}{I}) \right]$ + ($\alpha P_1+v+sP_1^2)T + C_{e_1}P_1T + C_{eh}h$ $I_1(\frac{e^{-\theta T}}{-\theta}+\frac{1}{\theta})+\frac{a(\gamma-1)}{I}(T+\frac{bT^2}{2}-\frac{bT}{I})$ + $C_t\left[I_1(\frac{e^{-\theta T}}{-\theta}+\frac{1}{\theta})+\frac{a(\gamma-1)}{I}(T+\frac{bT^2}{2}-\frac{bT}{I}) \right]$

Rhino brick industry

The differential equation to represent the production of rhino brick in the period $[0,T_1]$ is given by;

$$\frac{dB}{dt}+\phi B=a(\eta-1)e^{b_1 t} \tag{9.12}$$

Where $B(0)=B_1$

The inventory level of plastic in the factory during the period $[0,T_1]$ is given by

$$B(t)=B_1e^{-\phi t}+a(\eta-1)\frac{1}{\phi}(1+b_1t-\frac{b_1}{\phi})$$

Set up Cost (SC)= O_{rb}

Holding Cost(HC)= $h_{rb}\left[B_1(\frac{e^{-\phi T_1}}{-\phi}+\frac{1}{\phi})+\frac{a(\eta-1)}{\phi}(T_1+\frac{b_1T_1^2}{2}-\frac{b_1T_1}{\phi})\right]$

Deterioration Cost(DC) = $C_{rb}\ \phi\left[B_1(\frac{e^{-\phi T_1}}{-\phi}+\frac{1}{\phi})+\frac{a(\eta-1)}{\phi}(T_1+\frac{b_1T_1^2}{2}-\frac{b_1T_1}{\phi})\right]$

Production Cost(PC)= $(\alpha_{rb}P_2+v_{rb}+S_{rb}P_2^{\ 2})T_1$

Carbon Tax in Production (CT)= $C_{e_2}P_2T_1$

Transportation Cost of plastic from the dump yard(TC)= $\frac{1}{2}C_tP_2T_1$

Salvage Cost= $P\ \phi\left[B_1(\frac{e^{-\phi T_1}}{-\phi}+\frac{1}{\phi})+\frac{a(\eta-1)}{\phi}(T_1+\frac{b_1T_1^2}{2}-\frac{b_1T_1}{\phi})\right]$

Overall Cost of production of Rhino Brick is given by

OCRBPH(Overall Cost for Rhino Bricks Production House)= $O_{rb}+h_{rb}\left[B_1(\frac{e^{-\phi T_1}}{-\phi}+\frac{1}{\phi})+\frac{a(\eta-1)}{\phi}(T_1+\frac{b_1T_1^2}{2}-\frac{b_1T_1}{\phi})\right]+C_{rb}\ \phi$

$\left[B_1(\frac{e^{-\phi T_1}}{-\phi}+\frac{1}{\phi})+\frac{a(\eta-1)}{\phi}(T_1+\frac{b_1T_1^2}{2}-\frac{b_1T_1}{\phi})\right]+(\alpha_{rb}P_2+v_{rb}+S_{rb}P_2^{\ 2})T_1+C_{e_2}P_2T_1+$

$$\frac{1}{2}C_t P_2 T_1 + \mathrm{P}\phi \left[B_1\left(\frac{e^{-\phi T_1}}{-\phi}+\frac{1}{\phi}\right)+\frac{a(\eta-1)}{\phi}\left(T_1+\frac{b_1 T_1^2}{2}-\frac{b_1 T_1}{\phi}\right)\right]$$

Solution procedure

Total cost of the model is

$$\mathrm{TCF}=\mathrm{O}+ h\left[I_1\left(\frac{e^{-\theta T}}{-\theta}+\frac{1}{\theta}\right)+\frac{a(\gamma-1)}{I}\left(T+\frac{bT^2}{2}-\frac{bT}{I}\right)\right]$$

$$c\theta\left[I_1\left(\frac{e^{-\theta T}}{-\theta}+\frac{1}{\theta}\right)+\frac{a(\gamma-1)}{I}\left(T+\frac{bT^2}{2}-\frac{bT}{I}\right)\right]+ (\alpha P_1+v+sP_1^2)T + C_{e_1}P_1 T +$$

$$C_{eh} h \left[I_1\left(\frac{e^{-\theta T}}{-\theta}+\frac{1}{\theta}\right)+\frac{a(\gamma-1)}{I}\left(T+\frac{bT^2}{2}-\frac{bT}{I}\right)\right]+$$

$$C_t\left[I_1\left(\frac{e^{-\theta T}}{-\theta}+\frac{1}{\theta}\right)+\frac{a(\gamma-1)}{I}\left(T+\frac{bT^2}{2}-\frac{bT}{I}\right)\right]+ \mathrm{O}_{\mathrm{rb}}+$$

$$h_{rb}\left[B_1\left(\frac{e^{-\phi T_1}}{-\phi}+\frac{1}{\phi}\right)+\frac{a(\eta-1)}{\phi}\left(T_1+\frac{b_1 T_1^2}{2}-\frac{b_1 T_1}{\phi}\right)\right]+\mathrm{C_{rb}}\phi$$

$$\left[B_1\left(\frac{e^{-\phi T_1}}{-\phi}+\frac{1}{\phi}\right)+\frac{a(\eta-1)}{\phi}\left(T_1+\frac{b_1 T_1^2}{2}-\frac{b_1 T_1}{\phi}\right)\right]+(\alpha_{rb}P_2+v_{rb}+S_{rb}P_2^2)T_1 +$$

$$C_{e_2}P_2 T_1 + \frac{1}{2}C_t P_2 T_1 + \mathrm{P}\phi \left[B_1\left(\frac{e^{-\phi T_1}}{-\phi}+\frac{1}{\phi}\right)+\frac{a(\eta-1)}{\phi}\left(T_1+\frac{b_1 T_1^2}{2}-\frac{b_1 T_1}{\phi}\right)\right]$$

5. Numerical analysis

We put the following values of the variables and got the Global solution of the model. Values of the variables are as following.

O=1200,O_{rb}=1500, α=120,α_{rb}=100, v =1100, v_{rb} = 900, s= .006,s_{rb}=.002,C_{e1}=1.5, C_{e2}=0.8,C_{eh}=.75, h=0.5,h_{rb}=0.8, c=1.2,c_{rb}=1.8,c_t=2.85,p=6, ϕ =.002, θ =.007,I_1=350, B_1=200,T=29, a=2, r=1.2, I=320, B=180, γ =1.2, b=2, P_1=10,P_2=12, T_1=30, η =1.7, b_1=3

We got the values of the cost function as TCF= 137686.83.

6. Sensitivity analysis

We are changing are values of the variables in the model to see the behaviour of the total cost function. In this part, we have tried to find out the

role of different variables to the total cost function. We have changed the data firstly 50% increase, 50% decrease and 20% increase then 20% decreases. Then the corresponding changes in the cost function with the values of variables.

Table 1: This table shows the sensitivity of different parameters

Variables	Change of variable	Changed in cost function	% Changed in cost function
γ	1.8	172586.5	25.34% increase
	.8	172558.417	25.32% decrease
	1.2	137686.831	Model value
	1.44	172575.793	25.33% increase
	.96	172565.85	25.33% decrease
C_t	4.275	151905.575	10.32% increase
	1.425	125206.982	9% decrease
	2.85	137686.83	Model value
	3.42	143896.075	4.5% increase
	2.28	133216.6	3.24% decrease
V	1650	221675.65	60.99% increase
	550	158516.89	15.22% decrease
	1100	137686.831	Model value
	1320	204716.89	48.68% increase
	880	176245.65	28% decrease

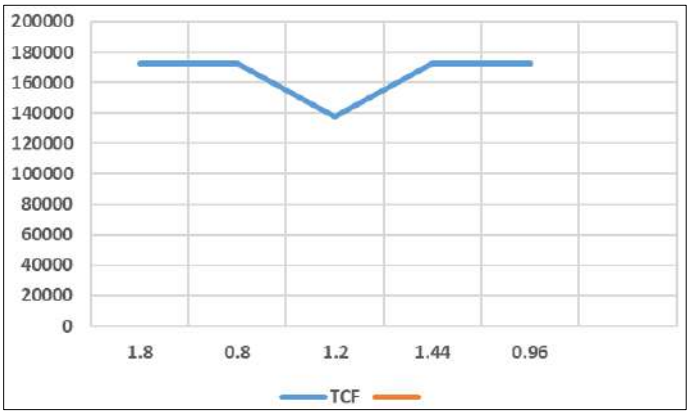


Figure 1: Depicts the relationship between γ and TCF

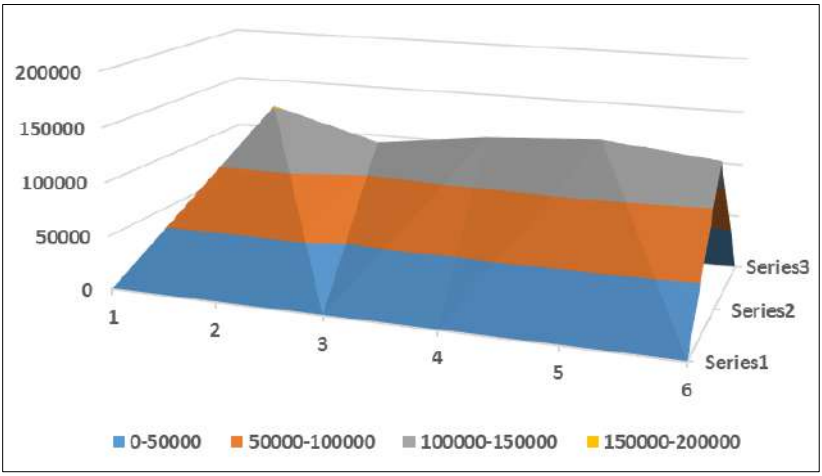


Figure 2: Depicts the relationship between C_t and TCF

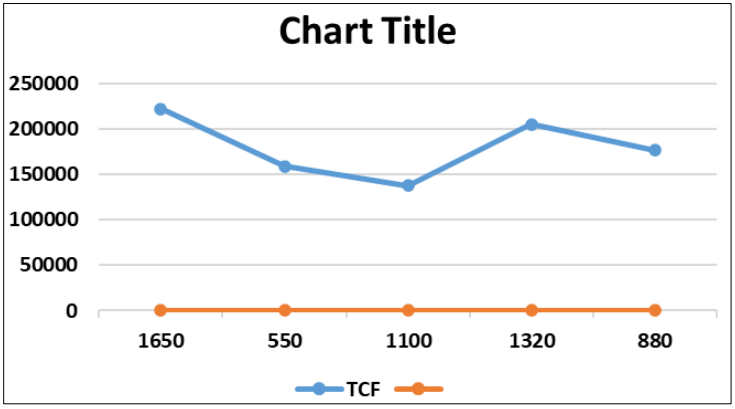


Figure 3: Depicts the relationship between v and TCF

Referrences

1. Sustainable Inventory Model with Carbon Emission Dependent Demand Under Different Carbon Emission Policies Soft Computing in Inventory Management, 2021, ISBN: 978-981-16-2155-0 Shikha Yadav, Farah Siddiqui, Aditi Khanna, DOI: 10.1007/978-981-16-2156-7_9
2. Chen, X., Benjaafar, S., & Elomri, A. (2013). The carbon-constrained EOQ. *Operations Research Letters, 41*, 172-179.

3. He, P., Zhang, W., Xu, X., & Bian, Y. (2015). Production lot-sizing and carbon emissions under cap-and-trade and carbon tax regulations. *Journal of Cleaner Production, 103*, 241-248.

4. Hota, S. K., Sarkar, B., & Ghosh, S. K. (2020). Effects of unequal lot size and variable transportation in unreliable supply chain management. *Mathematics, 8*, 357.

5. Li, L. (2016). Coordination contracts for two-echelon supply chain with price-and-carbon-emission dependent demand. *Modern Economy, 7*, 606.

6. Mishra, M., Hota, S. K., Ghosh, S. K., Sarkar, B. (2020a). Controlling waste and carbon emission for a sustainable closed-loop supply chain management under a cap-and-trade strategy. *Mathematics, 8*, 466.

7. Salehi, M., Jalalian, M., & Siar, M. M. V. (2017). Green transportation scheduling with speed control: Trade-off between total transportation cost and carbon emission. *Computers & Industrial Engineering, 113*, 392-404.

8. Tang, Z., Liu, X., Wang, Y., Ma, D. (2020). Integrated inventory-transportation scheduling with sustainability-dependent demand under carbon emission policies. *Discrete Dynamics in Nature and Society, 2020*.

9. Decarbonisation through production of Rhino Bricks from waste plastics: EPQ Model, 2022, N Bhattacharjee.et al., Inventory optimization, https://doi.org/10.1007/978-981-19-6337-7_9.

Chapter - 14

AR3D Face Recognition: A New Frontier in Human-Computer Interaction

Authors

Kamran Khan
Department of Computer Science and Engineering/Student, Swami Vivekananda University, Barrackpore, West Bengal, India

Biplab Saha
Department of Computer Science and Engineering/Student, Swami Vivekananda University, Barrackpore, West Bengal, India

Md Asif
Department of Computer Science and Engineering/Student, Swami Vivekananda University, Barrackpore, West Bengal, India

Anirban Das
Department of Computer Science and Engineering/Student, Swami Vivekananda University, Barrackpore, West Bengal, India

Pradip Sahoo
Department of Computer Science and Engineering/Student, Swami Vivekananda University, Barrackpore, West Bengal, India

Chapter - 14

AR3D Face Recognition: A New Frontier in Human-Computer Interaction

Kamran Khan, Biplab Saha, Md Asif, Anirban Das and Pradip Sahoo

Abstract

In an era marked by rapid technological innovation, Human-Computer Interaction (HCI) has evolved significantly, with the convergence of Augmented Reality (AR) and 3D Face Recognition technologies forming a new paradigm shift in HCI. AR3D Face Recognition is a transformative medium that merges the digital and physical realms, offering immersive experiences for users. It transcends the limitations of 2D recognition by capturing the rich depth and nuances of the human visage. This research paper delves into the core components of AR3D Face Recognition, addressing the intricate fusion of hardware and software that breathes life into this technology. It explores its applications across diverse domains, from revolutionizing the gaming and entertainment industry to its profound impact on healthcare, security, retail, and marketing. Each use case is a testament to the versatility and transformative potential of AR3D Face Recognition. However, such technological prowess comes with challenges. Privacy concerns, such as data security, consent, and potential misuse, demand careful consideration of data security, consent, and potential misuse. Technical obstacles, such as lighting conditions and algorithmic accuracy, necessitate innovative solutions. Ethical considerations, including bias mitigation and the preservation of individual rights, must guide our path forward. However, as we venture into this new world of AR3D Face Recognition, we face multifaceted challenges. Privacy concerns demand stringent safeguards for facial data and user consent, technical limitations call for creative solutions, and ethical considerations surrounding bias, discrimination, and consent necessitate vigilant oversight and principled development.

Keywords: Augmented reality, 3D face recognition, human-computer

interaction, facial biometrics, computer vision, emotion recognition, 3D facial reconstruction.

Introduction

Human-Computer Interaction (HCI) is a rapidly evolving field driven by technological innovation, with the convergence of Augmented Reality (AR) and 3D Face Recognition presenting a groundbreaking frontier known as AR3D Face Recognition. This fusion represents a significant step towards redefining human engagement with machines and digital environments. AR has evolved from science fiction into tangible reality, casting digital overlays upon our physical surroundings, while 3D Face Recognition technology has scaled unprecedented heights, transcending the limitations of conventional biometric identification by rendering the human face in three dimensions with astonishing precision and detail.

The union of AR and 3D Face Recognition within the framework of AR3D Face Recognition is an alliance of technology and humanity, offering a holistic approach to augmenting daily interactions with computers, devices, and digital content. This synergy opens new avenues for a wide spectrum of applications, from entertainment and healthcare to security and retail, each promising to revolutionize the way we engage with our environment, devices, and each other.

The transformative power of AR3D Face Recognition resonates across a multitude of domains, such as gaming and entertainment, healthcare, security and authentication systems, retail, and marketing. It reshapes gaming and entertainment by offering unparalleled immersion, enabling users to become active participants in digital narratives while their emotions and expressions shape the virtual world. In healthcare, it transcends the boundaries of patient identification, diagnostic accuracy, and remote telemedicine. In retail, consumers are greeted with tailored shopping experiences, and marketers glean insights from emotional responses, paving the way for personalized marketing strategies.

However, AR3D Face Recognition brings forth its share of challenges and ethical considerations. The preservation of privacy in an era of ubiquitous facial data collection, technical limitations related to environmental factors and device accessibility, and the imperative to address bias, discrimination, and consent issues demand our utmost attention and conscientious development.

The future promises to be a canvas of innovation and progress, with advances in AR hardware and the continual refinement of recognition algorithms holding the keys to unlocking hitherto unimagined applications and possibilities. The boundaries of AR3D Face Recognition will continue to expand, weaving itself into the tapestry of our lives and society in ways yet uncharted.

Background

Over the past five decades, augmented reality (AR) technology has profoundly transformed our interaction with the tangible world. The origins of AR trace back to the 1950s when Morton Heilig, a cinematographer, envisioned cinema as an immersive experience that would engage all the senses. In 1968, Ivan Sutherland pioneered AR by creating the first optical see-through head-mounted display system. Myron Krueger further pushed the boundaries in 1975 with the creation of Video Place, a room enabling users to interact with virtual objects.

In 1992, Louis Rosenburg, a researcher at the USAF Armstrong's Research Lab, introduced 'Virtual Fixtures,' marking a significant milestone in the evolution of AR technology. Julie Martin, a writer and producer, brought AR into the entertainment industry with the ground-breaking theatre production, "Dancing in Cyberspace," in 1994. The year 2000 witnessed the development of the first outdoor mobile AR game, AR Quake, by Bruce Thomas, showcased during the International Symposium on Wearable Computers.

In 2005, the Horizon Report anticipated the broader emergence of AR technologies within the following 4-5 years. Indeed, that year saw the development of camera systems capable of real-time environmental analysis and object-environment positional relationships. Subsequent years witnessed the proliferation of AR applications, particularly in the mobile domain, exemplified by the launch of the Wikitude AR Travel Guide in 2008 and the development of medical applications in 2007.

Google entered the AR scene in 2014 with the introduction of Google Glass, a pair of AR glasses offering immersive experiences. As society increasingly relies on mobile devices, the adoption of AR technology continues to surge. The future of AR lies in software advancements, given the ubiquity of smartphones among consumers, making it a convenient platform to deliver AR experiences to the masses.

Human-Computer Interaction (HCI) has evolved significantly since the inception of computing devices, with the pursuit of more natural, intuitive, and immersive ways for humans to interact with computers. Augmented Reality (AR) and 3D Face Recognition technologies, both influential in their own right, have merged to create AR3D Face Recognition, a transformative paradigm in HCI. AR3D Face Recognition is a result of the continuous evolution of HCI, driven by technological innovation and the pursuit of more immersive, intuitive, and responsive human behaviors and emotions.

Augmented Reality (AR) is a powerful technology that blurs the lines between the physical and digital worlds by overlaying virtual elements onto the real environment. This enhances human perception and interaction, opening up applications in gaming, education, and industry. Human-Computer Interaction (HCI) has evolved significantly since the inception of computing devices, with AR and 3D Face Recognition technologies merging to create AR3D Face Recognition. AR3D Face Recognition is a transformative paradigm in HCI, driven by technological innovation and the pursuit of more immersive, intuitive, and responsive human behaviors and emotions.

3D Face Recognition is a groundbreaking advancement in facial recognition technology, capturing and analyzing the intricate three-dimensional structure of a human face. This depth-aware approach allows for highly accurate identification, even in challenging lighting conditions or when facial expressions are in flux. This technology holds significant promise in security, healthcare, and personalization, offering a wealth of possibilities for applications such as tracking emotions, deciphering expressions, and enhancing user experiences.

AR3D Face Recognition is a technological revolution that merges the immersive capabilities of AR with the precision of 3D face recognition, transforming human-computer interaction (HCI) into a dynamic interface. This fusion allows users to interact with digital content in a more natural, empathetic, and contextually relevant manner, interpreting facial expressions, emotions, and identity. The transformative potential of AR3D Face Recognition is evident across various sectors, including gaming and entertainment, healthcare, retail, and marketing. In gaming and entertainment, AR3D Face Recognition allows users to become active participants in digital narratives, where their emotions dictate the story. In healthcare, AR3D Face Recognition streamlines patient identification and

diagnostics, facilitating telemedicine through secure and remote verification. Security and authentication systems benefit from foolproof identification methods, enhancing security in airports, borders, and surveillance contexts. In retail, AR3D Face Recognition fosters personalized shopping experiences and facilitates targeted marketing campaigns. However, as we explore this new frontier, we encounter multifaceted challenges. Privacy concerns demand robust safeguards for facial data and user consent. Technical limitations, such as environmental factors affecting 3D reconstruction and AR hardware accessibility, necessitate innovative solutions. Ethical considerations surrounding bias, discrimination, and consent require vigilant oversight and principled development. AR3D Face Recognition represents a significant advancement in HCI, combining the immersive capabilities of AR with the precision and depth of 3D face recognition. However, it also presents challenges such as privacy concerns, technical limitations, and ethical considerations.

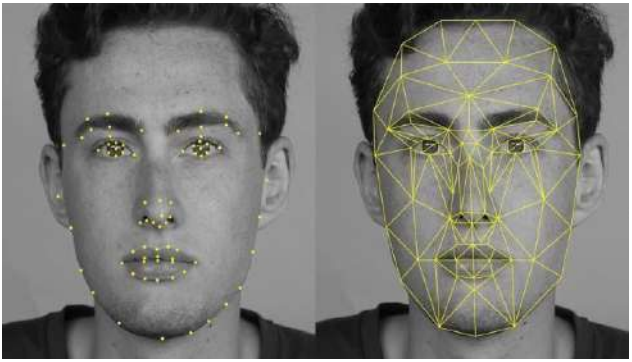
AR3D face recognition technologies

Augmented Reality (AR)

In the ever-evolving landscape of technology, Augmented Reality (AR) stands as a transformative frontier, weaving enchanting narratives where the physical and digital realms converge. It is a phenomenon that transcends the ordinary, imbuing our everyday experiences with a touch of the extraordinary. In this exploration, we embark on a journey through the captivating world of AR, unearthing its essence, potential, and the profound impact it wields on the fabric of our reality.

At its core, Augmented Reality is an intricate dance, choreographed meticulously between the tangible world that envelops us and the intangible realm of digital information. AR transcends the constraints of traditional interfaces, allowing us to see, hear, and even feel the invisible, superimposing computergenerated content onto our perception of reality. This harmonious blend catalyses an immersive multisensory experience, redefining how we interact with the digital tapestry of the universe.

The roots of AR extend deep into the annals of computer science history. Ivan Sutherland's "Sword of Damocles" in the 1960s laid the foundation, albeit in a nascent form. Decades later, technological advancements in processing power, miniaturization, and sensory capabilities breathed life into AR's full potential. What was once science fiction has now become a tangible part of our lives.



Augmented reality: The symphony of key components and technologies in the enchanting realm of Augmented Reality (AR), the digital and physical harmonize to create immersive, multi-sensory experiences. To unveil the magic behind AR, we must delve into the symphony of key components and technologies orchestrating this delicate fusion of worlds. Each element plays a unique note in this digital symphony, forging a path towards augmented enlightenment.

Sensor emissaries: Weaving reality into data

AR's journey commences with a chorus of sensors that extend our perception. These emissaries of the digital realm include:

- **Cameras:** The eyes of AR, cameras capture the physical world, providing the canvas upon which digital overlays are painted. Depth-sensing cameras, like LiDAR, add an extra dimension, enabling more precise object placement.
- **Gyroscope and accelerometer:** These gyroscopic wizards and accelerative artisans provide orientation and movement data, ensuring that digital content aligns seamlessly with our perspective.
- **GPS and location sensors:** Geolocation technologies anchor AR experiences to specific places, enhancing context-aware applications, such as navigation and location-based gaming.

Visionaries of the virtual: Algorithms and computer vision

AR's true magic lies in its ability to understand and interact with the physical world. Computer vision and algorithms are the visionaries, interpreting sensor data and making sense of our surroundings:

- **Object recognition:** These algorithms identify and track objects,

enabling AR to interact with them. From recognizing your coffee mug to tracking constellations in the night sky, object recognition paints the canvas with context.

- **Simultaneous Localization and Mapping (SLAM):** SLAM algorithms, like cartographers of the digital realm, build maps of the environment in real time, allowing AR to navigate and place digital elements with precision.

Portals to the virtual: Display technologies

AR's canvas, or rather, window into the digital realm is presented through a variety of technologies, each offering a unique perspective:

- **Head-Mounted Displays (HMDs):** Devices like Microsoft's HoloLens and Magic Leap provide immersive, hands-free AR experiences. These headsets project digital content directly into your field of view, seamlessly blending with the real world.
- **Smartphones and tablets:** Everyday devices equipped with AR capabilities serve as accessible portals. By leveraging their cameras and processing power, they can overlay digital information onto the physical environment.
- **Wearable glasses:** These stylish companions, like Google Glass, offer a less intrusive means of experiencing AR, placing digital information in your peripheral vision.

The artistry of content creation: Development tools

AR's magic is crafted by content creators who use specialized tools to breathe life into the digital overlays:

- **Unity 3D:** This versatile development platform empowers creators to craft 3D models, animations, and interactive elements that seamlessly blend with the real world.
- **AR Development Kits (ARKit, ARCore):** Provided by tech giants like Apple and Google, these development kits offer a treasure trove of tools and libraries for building AR applications tailored to their respective ecosystems.

Examples of popular AR platforms and devices

The AR landscape is dynamic, with a constellation of platforms and devices vying for attention. Some notable examples include:

1. **Microsoft HoloLens:** A pioneer in the mixed reality domain, HoloLens offers a wearable AR headset that blends holographic digital content with the real world. It has found applications in fields ranging from architecture to healthcare.
2. **Apple ARKit:** As part of Apple's commitment to AR, ARKit provides developers with robust tools to create AR experiences on iOS devices. Apps like Pokemon Go and IKEA Place exemplify its potential.
3. **Google ARCore:** Similar to ARKit but for Android, ARCore powers a growing ecosystem of AR apps and experiences on a wide range of Android devices.
4. **Snapchat and instagram filters:** These popular social media platforms have integrated AR features, allowing users to overlay filters, animations, and effects onto their photos and videos, making AR part of everyday communication.

Applications of AR3D face recognition

AR3D Face Recognition, the captivating union of Augmented Reality (AR) and 3D face recognition technologies, unveils a rich tapestry of applications that transcend the conventional boundaries of human-computer interaction. As we step into this transformative realm, we are met with a dazzling array of possibilities that reshape industries, enhance experiences, and redefine the way we connect with technology and each other.

Gaming and entertainment: Crafting immersive realms

In the realm of gaming and entertainment, AR3D Face Recognition emerges as a luminary, weaving immersive narratives where users' expressions and emotions become the conduits of engagement:

- **Immersive gaming experiences:** Game characters come alive, mirroring players' expressions and emotions, creating deeply engaging and personalized adventures.
- **Real-time emotion recognition:** AR3D Face Recognition captures emotional responses, allowing games to adapt dynamically, leading to exhilarating and unpredictable gameplay.
- **Interactive storytelling:** In interactive narratives, characters react to users' expressions, forging an emotional connection that transcends traditional storytelling.

Healthcare: A guardian of well-being

In healthcare, AR3D Face Recognition dons the mantle of guardianship, ensuring patient identification and diagnostic precision:

- **Patient identification and verification:** It provides secure and frictionless patient identification, safeguarding medical records and preventing identity errors.
- **Medical diagnostics:** Healthcare professionals employ AR3D Face Recognition for noninvasive scans, aiding in early diagnosis and treatment planning.
- **Mental health monitoring:** Emotion analysis assists therapists in assessing patients' mental states, facilitating personalized mental health care.

Security and authentication: Fortifying access control

In the realm of security and authentication, AR3D Face Recognition emerges as a sentinel, safeguarding access:

- **Secure access control systems:** AR3D Face Recognition enhances access security by ensuring that only authorized individuals gain entry to restricted areas.
- **Border control and airport security:** It strengthens border control and airport security, facilitating quick and secure identity verification.
- **Surveillance and crime prevention:** Law enforcement agencies employ this technology for real-time tracking and identification, aiding in crime prevention and investigation.

Retail and marketing: Personalization and engagement

In the retail and marketing landscape, AR3D Face Recognition becomes a virtuoso, personalizing experiences and captivating customers:

- **Personalized shopping experiences:** Retailers leverage AR3D Face Recognition to offer personalized product recommendations and virtual try-ons.
- **Targeted advertising:** Emotion-aware advertising adjusts content based on viewers' reactions, enhancing engagement and ad effectiveness.

- **Customer engagement and analytics:** Analyzing facial expressions and responses provides insights into customer preferences and behaviors, allowing businesses to refine their strategies.

Challenges and limitations of AR3D face recognition

As Augmented Reality (AR) and 3D face recognition converge into the captivating realm of AR3D Face Recognition, they bring with them a Pandora's box of opportunities and complexities. While the promises are boundless, the path forward is not without its thorns. In this exploration, we shed light on the intricate challenges and limitations that accompany this revolutionary technology.

Privacy concerns: Guarding the sanctity of personal data

One of the foremost concerns surrounding AR3D Face Recognition is the preservation of privacy. The technology involves the capture and processing of facial data, raising questions about consent, data security, and the potential for misuse. Striking a balance between innovation and safeguarding individuals' privacy remains a formidable challenge.

Data security and privacy

Storing and transmitting facial data securely is a paramount concern. Breaches or unauthorized access to this sensitive information can lead to identity theft and privacy violations.

Consent and user control

Ensuring that users are aware of and consent to the collection and use of their facial data is essential.

Providing individuals with control over how their data is used and stored is equally critical.

Potential misuse

The technology can be exploited for surveillance, tracking, or even impersonation if not carefully regulated and monitored.

Technical challenges: Illuminating the shadows

AR3D Face Recognition relies on a delicate interplay of hardware and software, presenting several technical hurdles:

Lighting and environmental factors

Variability in lighting conditions, such as low light or harsh sunlight, can impact the accuracy of 3D face reconstruction and recognition, making robust performance in all environments a challenge.

Hardware limitations

The effectiveness of AR3D Face Recognition heavily depends on the capabilities of the hardware. Affordable consumer-grade devices may lack the necessary sensors and processing power for seamless performance.

Algorithmic accuracy and robustness

Achieving high accuracy in recognizing faces across diverse facial expressions, ages, and ethnicities remains an ongoing challenge. Reducing bias in recognition algorithms is crucial to ensure fairness.

Bias and fairness

Recognition algorithms may exhibit bias, leading to inaccurate results for certain demographic groups. Eliminating bias and ensuring fairness in recognition is a complex ethical endeavor.

Discrimination issues

AR3D Face Recognition could exacerbate existing societal biases and discrimination, particularly in law enforcement and surveillance applications.

Pioneering future directions for AR3D face recognition

As we stand on the precipice of possibility, AR3D Face Recognition beckons us towards a future that defies conventional boundaries in human-computer interaction. This synergy of Augmented Reality and 3D face recognition holds within it the seeds of innovation that will reshape industries, revolutionize experiences, and redefine our very notion of connection with technology. As we peer into the horizon, let us embark on a journey of exploration into the promising future directions this transformative technology may unveil.

Advancements in AR hardware: Shaping immersive experiences

The evolution of AR hardware is poised to usher in an era of even more immersive experiences. Envision lightweight, stylish AR glasses with extended battery life, offering wearers a seamless blend of digital and

physical realities. This convergence of style, functionality, and affordability will make AR accessible to a broader audience, amplifying its impact in domains ranging from education to entertainment.

Beyond entertainment: Emerging applications

While AR3D Face Recognition has already begun to revolutionize gaming and entertainment, its tendrils will extend into previously uncharted territories:

- **Healthcare augmentation:** AR3D Face Recognition will play a pivotal role in telemedicine, aiding healthcare providers in accurately assessing patients' conditions through remote facial diagnostics.
- **Education revolution:** The technology will transform classrooms into immersive learning spaces, where teachers gauge students' engagement and emotions in real time, adapting lessons accordingly.
- **Remote work and collaboration:** AR-enhanced video conferencing will provide new dimensions of virtual presence, with facial recognition tracking ensuring that online interactions feel as personal as in-person meetings.

Wearable augmented reality in daily life

Imagine AR glasses as a ubiquitous accessory, akin to today's smartphones. From real-time language translation during travel to step-by-step augmented cooking tutorials in your kitchen, these wearables will become indispensable tools in our daily lives.

Enhanced emotional understanding

AR3D Face Recognition will evolve to read emotions with higher accuracy and nuance. It will facilitate not only personalized content recommendations but also mental health support through real-time emotional analysis, offering insights and interventions when needed.

Cross-platform compatibility

Efforts will be made to standardize AR3D Face Recognition technologies, enabling cross-platform compatibility. This interoperability will foster a richer ecosystem of AR applications, benefiting both developers and users.

Augmented reality in the web

AR will transcend app-based experiences, seamlessly integrating into web browsers. This democratization of AR content will make it more accessible, allowing users to interact with augmented elements directly from their browsers, unlocking a new dimension of digital engagement.

Conclusion

In the mosaic of human-computer interaction, AR3D Face Recognition stands as a brilliant, transformative stroke of artistry. Its fusion of Augmented Reality and 3D face recognition technologies paints a canvas filled with possibilities that challenge the boundaries of imagination. As we draw this exploration to a close, we find ourselves at a crossroads where technology and humanity intertwine, beckoning us towards a future both exhilarating and nuanced. AR3D Face Recognition is not merely a technological innovation; it is an ode to our innate human desire for connection, understanding, and engagement. It redefines the contours of how we interact with the digital universe, ushering in an era where our expressions and emotions bridge the chasm between the physical and the virtual. It is an invitation to explore a digital frontier where innovation and ethics must walk hand in hand. However, with great innovation comes great responsibility. AR3D Face Recognition is not without its challenges, be they privacy concerns, technical complexities, or ethical dilemmas. The responsible development and ethical deployment of this technology will be our litmus test, where we must tread carefully to ensure that the promises it holds are not overshadowed by unintended consequences. As we move forward into the uncharted territory that AR3D Face Recognition illuminates, let us carry with us a commitment to safeguarding privacy, addressing bias, and respecting consent. Let us forge a future where technology enriches our lives while preserving our fundamental human rights and dignity. In the convergence of Augmented Reality and 3D face recognition, we find a harmonious symphony of possibilities, where innovation and ethics dance in a delicate balance. It is a future where our digital interactions are as authentic and profound as our physical ones, where technology enhances rather than diminishes our humanity. AR3D Face Recognition is an invitation to dream, to create, and to explore, and in answering that call with responsibility and vision, we embark on a journey towards a more connected, enriched, and harmonious digital future.

References

1. Academic Databases: Search in databases such as IEEE Xplore, ACM Digital Library, Google Scholar, PubMed, or other specialized research databases. Use keywords like "AR3D Face Recognition," "Augmented Reality 3D Face Recognition," or related terms to narrow down your search.
2. Research Papers: Look for peer-reviewed research papers, conference proceedings, and journal articles related to AR3D Face Recognition. Pay attention to recent publications to ensure you have the latest information.
3. Authors and Experts: Identify researchers, authors, or experts in the field of AR, facial recognition, and computer vision who have published relevant work. Their papers can be valuable sources for your research.
4. Citations: Check the citations in relevant papers to discover additional sources that might be beneficial for your research.
5. Reputable Journals and Conferences: Seek publications in reputable journals and conferences related to computer vision, augmented reality, and human-computer interaction.
6. Books and Reports: Look for books, reports, or whitepapers that provide in-depth information on the topic.
7. L. CY, M. Shpitalni, and R. Gadh, "Virtual an augmented reality technologies for product realization," CIRP Annals 199: Manufacturing Technology, vol. 48, pp. 471-495, 1999.
8. R.T. Azuma, "A survey of Augmented Reality," In Presence: Teleoperator and Virtual Environment, pp. 355-385, 1997.
9. Yunqiang Chen, Qing Wang, Hong Chen, Xiaoyu Song, Hui Tang, Mengxiao Tian, “An overview of augmented reality technology”.

Chapter - 15

Augmented Reality and Virtual Reality in Education: A Transformative Journey into Immersive Learning Environments

Authors

Md Asif
Department of Computer Science and Engineering/Student, Swami Vivekananda University, Barrackpore, West Bengal, India

Aneek Mondal
Department of Computer Science and Engineering/Student, Swami Vivekananda University, Barrackpore, West Bengal, India

Sarvesh Soumil
Department of Computer Science and Engineering/Student, Swami Vivekananda University, Barrackpore, West Bengal, India

Anirban Das
Department of Computer Science and Engineering/Student, Swami Vivekananda University, Barrackpore, West Bengal, India

Pradip Sahoo
Department of Computer Science and Engineering/Student, Swami Vivekananda University, Barrackpore, West Bengal, India

Chapter - 15

Augmented Reality and Virtual Reality in Education: A Transformative Journey into Immersive Learning Environments

Md Asif, Aneek Mondal, Sarvesh Soumil, Anirban Das and Pradip Sahoo

Abstract

In recent years, Augmented Reality (AR) and Virtual Reality (VR) technologies have gained significant traction in education, presenting a dynamic and transformative landscape. This research paper explores the theoretical underpinnings, practical applications, and potential benefits and challenges of incorporating AR and VR into educational contexts. The interdisciplinary approach considers their impact on teaching methods, student engagement, and the development of 21st-century skills. Emerging trends, ethical concerns, and the future potential of AR and VR in shaping education are also discussed. In essence, this paper underscores how the integration of AR and VR technologies in education is reshaping the learning experience for both students and educators. It offers insights into the multifaceted dimensions of AR and VR, encompassing pedagogy, cognition, and practical implementation. The aim is to provide valuable insights for educators, researchers, and policymakers navigating the evolving landscape of immersive educational technologies. The paper emphasizes that AR and VR have the potential to revolutionize education by enhancing engagement, improving outcomes, and creating interactive and captivating learning environments.

Keywords: Augmented Reality (AR), Virtual Reality (VR), experiential learning, customized learning.

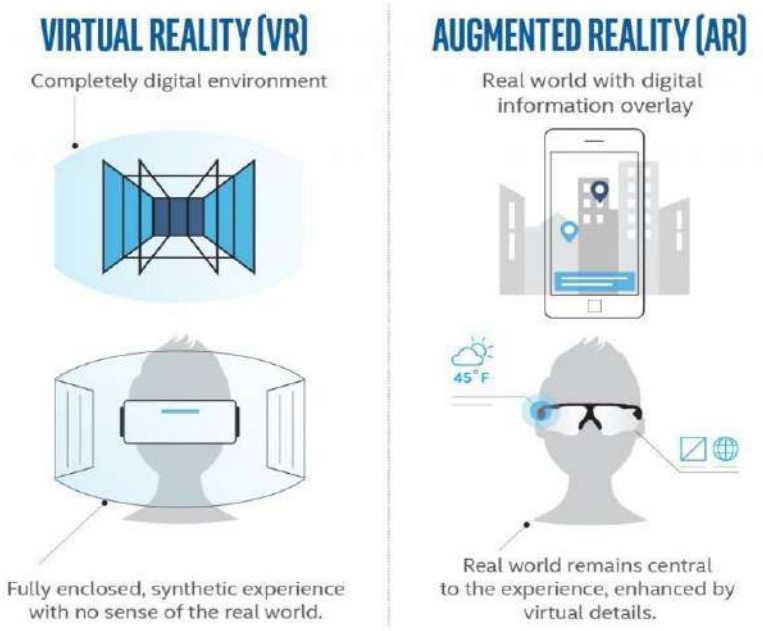
Introduction

In an age marked by rapid technological advancements, Augmented Reality (AR) and Virtual Reality (VR) have emerged as transformative forces that transcend their initial associations with entertainment and gaming

(Klopfer, 2019; Milgram & Kishino, 1994). These immersive technologies have evolved significantly, transitioning from the realm of science fiction to tangible tools that are reshaping our daily experiences (Sutherland, 1968). Within this context, gaining insight into the background and context of AR and VR is crucial to grasp the depth of their impact and their potential to redefine various aspects of our lives.

The journey of AR and VR from speculative concepts to real-world applications has been marked by remarkable technological progress, increased accessibility, and innovative use cases (Biocca, 1992; Steuer, 1992). These technologies have not only revolutionized entertainment and gaming but have also found substantial utility in education, training, and therapeutic interventions (Dede, 2009; Rizzo *et al*., 2001). Of particular interest is the profound impact of AR and VR in education, offering an exciting avenue for exploration (Wu *et al*., 2013).

In this paper, we embark on a comprehensive exploration of Augmented Reality (AR) and Virtual Reality (VR) in education. We begin by tracing the historical development of these technologies, offering valuable insights into their journey from conceptual origins to their current state of technological prowess (Azuma, 1997; Kruijff *et al*., 2010). We then delve into the diverse range of current applications of AR and VR in educational settings, highlighting their integration into various subjects and levels of education, from immersive learning experiences to specialized training (Merchant *et al*., 2014; Pivec, 2008). Additionally, we examine the advantages and challenges associated with their use, shedding light on improved engagement, personalized learning experiences, and potential considerations such as cost and accessibility (Dunleavy *et al*., 2009; Lee & Wong, 2014). Looking forward, we discuss the promising future prospects of AR and VR in education, including emerging technologies like Mixed Reality (MR) (Milgram & Colquhoun, 1999). Finally, we touch upon the implications for educational policy and practice, emphasizing the importance of ethical integration and continued research and development in this dynamic field (Carmigniani *et al*., 2011; Dede, 2016).



Demystifying the world of virtual and augmented realities

Virtual Reality (VR) and Augmented Reality (AR) are two immersive technologies that have captured the world's imagination due to their transformative potential in altering our perception of reality. Despite their distinct experiences, these technologies share common historical origins and have undergone remarkable evolution over the years. The concept of immersing individuals in simulated environments has a surprisingly long history, dating back to the 19th century. Sir Charles Wheatstone's introduction of the "stereoscope" in 1838, a device capable of creating a three-dimensional illusion by presenting distinct images to each eye, laid the initial groundwork for the fundamental principles that underpin both VR and AR.

The history of Virtual Reality may appear rooted in the digital age, but its origins can be traced back to the mid-20th century, suggesting that immersive, simulated experiences predate our modern technological landscape. Although the term "Virtual Reality" was not coined until the late 1980s by Jaron Lanier, the conceptual groundwork for VR began much earlier. Pioneers in the 1950s and 1960s laid the initial conceptual bricks for what would eventually become VR. During this period, rudimentary flight

simulators aimed at providing realistic experiences for pilots and astronauts marked the initial exploration of immersive technology. However, these early systems were a far cry from the immersive VR environments we know today. A pivotal moment in the history of VR occurred in 1968 when Ivan Sutherland, often referred to as the father of computer graphics, introduced the world to the "Sword of Damocles." This invention represented one of the earliest head-mounted displays (HMDs) and demonstrated the groundbreaking concept of tracking head movements to update the displayed scene accordingly. This innovation laid the groundwork for a fundamental aspect of modern VR experiences. The 1980s saw the emergence of commercial VR systems, marking a crucial chapter in the technology's evolution. Innovations like the VPL Research Data Glove and the Eyephone HMD introduced the world to the potential of VR. However, it's important to note that during this period, VR technology was still in its nascent stages, and the prohibitive cost of these systems limited their widespread adoption.

The 1990s and early 2000s witnessed VR gaining momentum, particularly within the gaming industry. Releases like Nintendo's Virtual Boy and various arcade VR experiences attempted to bring VR to the masses. Nevertheless, these early endeavors faced limitations, including issues with graphics quality, motion sickness, and accessibility, preventing VR from reaching its full potential. The true resurgence of VR occurred in the 2010s, marked by a convergence of factors that propelled the technology into the mainstream. First, advancements in powerful and affordable graphics processing units (GPUs) paved the way for the creation of immersive 3D environments boasting realistic graphics. Simultaneously, innovations in motion tracking and the introduction of head-mounted displays (HMDs) like the Oculus Rift, HTC Vive, and PlayStation VR brought VR experiences closer to the broader public.

Crucially, the gaming industry played a pivotal role in the popularization of VR during this era. Games such as "Half-Life: Alyx" and "Beat Saber" not only showcased the potential of VR but also cultivated a dedicated user base drawn to the immersive worlds and novel gameplay experiences VR offered. Beyond gaming, VR found applications in various sectors, including healthcare, where it was employed for pain management, therapy, and surgical training. Architects and designers harnessed VR to create and explore 3D models of buildings and products. Additionally, the entertainment industry explored VR for cinematic experiences and storytelling, pushing the boundaries of immersive narratives. This

multifaceted growth solidified VR as a transformative technology with the potential to revolutionize diverse aspects of human life.

Augmented Reality (AR) follows a parallel but distinct path from VR. While VR immerses users in entirely digital environments, AR overlays digital content onto the real world. The concept of augmenting reality with digital information dates back to the 1960s, with computer scientist Ivan Sutherland playing a notable role by creating the first AR head-mounted display known as the "Sword of Damocles." However, AR truly gained prominence in the 21st century with the advent of smartphones equipped with cameras and sensors. This technological leap allowed for the creation of AR applications, including early games and interactive marketing experiences.

Nevertheless, it wasn't until 2016 that AR made a significant impact on mainstream culture with the release of "Pokémon GO." This mobile game utilized AR technology to place virtual creatures within the real world via smartphones, captivating millions of users worldwide and showcasing AR's potential for entertainment and social interaction. The evolution of AR continued into the 2020s, marked by the emergence of AR headsets such as Microsoft's HoloLens and the Magic Leap One. These devices represent the next step in AR's development, blending augmented digital content seamlessly with the physical world, thereby opening up new possibilities in gaming, design, and workplace applications.

In conclusion, the histories of Virtual Reality and Augmented Reality are narratives of innovation, perseverance, and technological evolution. From their conceptual beginnings to their current prominence, these immersive technologies have undergone remarkable transformations and have the potential to reshape various aspects of human life. As technology continues to advance and societal acceptance deepens, VR and AR are poised to become integral parts of our daily lives, forever changing how we interact with the digital and physical worlds.

Pedagogical benefits

Enhancing engagement

In the realm of education, Augmented Reality (AR) and Virtual Reality (VR) seamlessly merge with reality, sparking innovation and captivating learners. AR and VR redefine learning in the digital age by transcending traditional pedagogy. They transform mundane classrooms into extraordinary experiences, where textbooks come to life and history becomes

an immersive journey through time. In VR, the boundaries of reality vanish as students explore diverse environments, kindling motivation through curiosity and engagement. These technologies empower learners to be active participants, driving their own education by interacting with digital elements. Intrinsic motivation flourishes as students uncover hidden knowledge and embark on quests fueled by imagination and intellect. Beyond captivation and motivation, AR and VR inspire wonder and curiosity, igniting a lifelong love for learning. These technologies weave a tapestry of engagement, motivation, and inspiration, expanding the boundaries of what's possible in education and nurturing the magic of learning.

Fostering a culture of engagement and curiosity is like orchestrating an intellectual symphony where active participation and inquisitiveness are encouraged. It involves nurturing the spark of human intellect, kindling a vibrant exchange of ideas, and breaking free from conventional thinking. Picture a diverse tapestry of minds, each contributing a unique thread of thought to create an enlightening mosaic. Curiosity, like a shuttlecock, weaves through this tapestry, connecting threads and forming patterns of uncharted knowledge. Promoting active participation is like breathing life into a dormant garden, nurturing engagement in fertile minds. It's about encouraging minds to not just absorb knowledge but actively explore, dissect, and understand it. This approach transforms learners into intellectual acrobats, captivating audiences with their mental agility. Curiosity, an unquenchable thirst for the unknown, guides us through life's maze, propelling us to seek answers to unasked questions. To foster curiosity is to ignite the flame of inquiry, urging the pursuit of undiscovered knowledge. In this tapestry, educators serve as gentle weavers, guiding without imposing, presenting the world as an unexplored museum of wonders. Learners become alchemists, transforming information into wisdom through active exploration. Curiosity takes center stage in this symphony of intellect, with active participation as its rhythm. Together, they create a harmonious masterpiece, where questions and answers blend seamlessly, captivating a growing community of seekers. Championing active participation and curiosity nourishes human intellect, turning minds into perpetual fountains of wonder. This unique tapestry of thought embodies the essence of human potential, where the journey of discovery becomes an exquisite art form.

Personalized learning

In the dynamic landscape of education, a ground-breaking shift is occurring, transcending conventional teaching methods, and ushering in the

era of Adaptive Learning Paths in immersive environments, where Augmented Reality (AR) and Virtual Reality (VR) combine to create customized learning experiences. This digital transformation accommodates various learning styles by seamlessly merging the tangible and intangible aspects of knowledge acquisition. Learners experience tailored education as the environment adapts to their preferences, catering to visual, auditory, or tactile learners. The adaptive system tracks progress and transforms challenges into opportunities for growth. This synergy between AR and VR reshapes traditional learning, offering a future where education is a personalized journey that celebrates diversity and responds to individual curiosity and understanding.

Medical education

In the field of medicine, VR and AR are transforming medical training and practice. Medical students can use VR for realistic surgical simulations, refining their skills in a safe environment before actual procedures [Reference: "Validation of the Osso VR Surgical Training Platform: A Randomized Controlled Trial" (Journal of Shoulder and Elbow Surgery, 2020)]. Surgeons employ AR for enhanced visualization during complex surgeries, overlaying vital patient data onto their view for precision [Reference: "Mixed Reality for Orthopedic Surgical Visualization" (Journal of Bone and Joint Surgery,

2019)]. VR aids patient engagement and rehabilitation, offering pain distraction and therapy exercises [Reference: "Virtual Reality for the Assessment of Frontal Lobe Dysfunction" (Journal of Neurological Sciences, 2018)]. Additionally, AR and VR have been crucial in telemedicine during the pandemic, enabling remote consultations and improving patient access to healthcare. They also play a role in pain management and stress reduction, providing patients with immersive experiences to alleviate discomfort [Reference: "Virtual Reality for the Management of Pain in Hospitalized Patients: Results of a Controlled Trial" (Journal of Pain and Symptom Management, 2017)].

Engineering

AR and VR technologies have a profound impact on various aspects of engineering. Engineers utilize AR and VR to visualize 3D designs in real-time, detect design flaws, optimize components, and assess ergonomics, thus saving time and resources during product development [Reference: "Virtual

Reality in Engineering: A State-of-the-Art Review" (IEEE Access, 2020)]. Additionally, these technologies enable engineers to simulate and test complex scenarios, providing valuable insights for design refinement before construction begins. AR enhances maintenance and performance optimization by overlaying real-time sensor data on physical equipment [Reference: "The Application of Virtual Reality in the Manufacturing Industry: A Review of Its Potentials" (Procedia CIRP, 2016)]. Moreover, AR and VR facilitate collaborative engineering, reducing the need for physical travel and enhancing efficiency [Reference: "Collaborative Virtual Environments in Industrial Design Engineering: A Review" (Computers in Industry, 2018)]. Furthermore, they are instrumental in training the next generation of engineers, offering immersive, hands-on learning experiences to bridge the gap between theoretical knowledge and practical application [Reference: "Virtual Reality in Education: A Tool for Learning in the Experience Age" (Frontiers in Robotics and AI, 2021)]. Lastly, in maintenance engineering, AR enhances efficiency and accuracy by providing real-time, context-sensitive information through smart glasses or devices [Reference: "Augmented Reality in Industrial Maintenance: An Overview" (Robotics and Computer-Integrated Manufacturing, 2021)].

Science

AR and VR technologies have transformed the scientific landscape in several key ways. They enable scientists to immerse themselves in otherwise inaccessible environments, facilitating exploration and visualization. For instance, marine biologists use VR to study ocean depths, while astronomers explore celestial bodies, providing novel perspectives [Reference: "Virtual Reality in Space Science Education and Outreach" (Acta Astronautica, 2020)]. These technologies are also invaluable for data visualization, aiding scientists in comprehending complex data, especially in genomics and molecular biology [Reference: "Virtual Reality and Augmented Reality for Molecular Sciences" (Trends in Biotechnology, 2018)]. VR aids in simulation and training, allowing researchers to conduct experiments and practice procedures without risk [Reference: "Immersive Virtual Reality in the Chemistry Laboratory: Can It Be Used to Enhance Student Learning?" (Journal of Chemical Education, 2018)]. Moreover, AR and VR revolutionize science education, offering immersive learning experiences that make subjects engaging and accessible to students worldwide [Reference: "Enhancing Science Education Through Immersive Virtual Field Trips" (Computers & Education, 2019)]. Additionally, they foster global

research collaboration by enabling scientists to meet virtually, transcending geographical barriers and enhancing the pace of scientific discovery [Reference: "Virtual Reality in Collaborative Research Environments: A Review" (Virtual Reality, 2021)].

History and social science

AR and VR are transforming historical and social sciences fields. They enable historical reenactments, offering students immersive experiences of past events, as seen in projects like "Rome Reborn" [Reference: "Rome Reborn 2.2: A Tour of Ancient Rome in 320 CE" (Journal of Cultural Heritage, 2013)]. Museums enhance exhibits using AR, superimposing context onto artifacts, turning museums into interactive learning environments [Reference: "Augmented Reality in Museums: Designing Contextual Experiences" (Museum Management and Curatorship, 2019)]. VR aids archaeological exploration, reconstructing ancient sites and revolutionizing the field [Reference: "Virtual Reality Applications in Archaeology" (Archaeological Prospection, 2020)]. Social scientists utilize VR for simulations, shedding light on human behavior [Reference: "The Promise of Virtual Reality in Behavioral Research and Therapy" (International Journal of Cognitive Therapy, 2019)]. Furthermore, AR and VR play a vital role in cultural preservation by digitizing artifacts and enabling remote collaboration on restoration [Reference: "Augmented Reality for Cultural Heritage: Foundations and Case Studies" (Journal on Computing and Cultural Heritage, 2016)]. These technologies bring history and social sciences to life, offering immersive experiences, enriching research, and preserving cultural heritage.

Distance learning

AR and VR technologies are enriching distance learning by creating virtual classrooms that foster interaction and collaboration among students and instructors, as well as offering immersive field trips, practical training, and language learning experiences, enhancing vocabulary acquisition and cultural immersion. These technologies enable complex simulations across various fields of study, from medicine to engineering, providing hands-on skills development and facilitating access to expert instruction from around the world in real-time, thereby fostering inclusivity and customization for diverse learning needs [References: "Virtual Reality in Education: A Tool for Learning in the Experience Age" (Frontiers in Robotics and AI, 2021); "Immersive Learning: A Survey of Its Status, Challenges, and Future

Trends" (IEEE Transactions on Emerging Topics in Computing, 2020); "Virtual Reality and Language Learning" (Cambridge University Press, 2020); "Virtual Reality Simulators in Medicine: A Scoping Review and Meta-Analysis" (Studies in Health Technology and Informatics, 2020); "Immersive Technologies for Learning: A Critical Review of the Learning Sciences Literature" (Educational Technology Research and Development, 2020); "Augmented Reality in Special Education: A Systematic Review" (Journal of Special Education Technology, 2018)].

Advantages of augmented reality & virtual reality

Augmented Reality (AR) and Virtual Reality (VR) offer numerous advantages in the field of education, transforming traditional teaching methods and enhancing the learning experience. Here are some key advantages,

Enhanced engagement: AR and VR make learning more interactive and engaging. Students can explore three-dimensional objects, environments, and simulations, which capture their attention and increase motivation to learn [Reference: "Immersive Learning: A Survey of Its Status, Challenges, and Future Trends" (IEEE Transactions on Emerging Topics in Computing, 2020)].

Experiential learning: These technologies enable experiential learning by immersing students in realistic scenarios. This hands-on approach allows them to actively participate in their education, leading to deeper understanding and retention of information [Reference: "Virtual Reality in Education: A Tool for Learning in the Experience Age" (Frontiers in Robotics and AI, 2021)].

Customized learning: AR and VR can adapt content to individual learning styles and paces. Students can progress at their own speed, receive personalized feedback, and access supplementary materials, tailoring the learning experience to their needs [Reference: "Personalized Learning: A Guide for Engaging Students with Technology" (International Society for Technology in Education, 2017)].

Accessibility: These technologies can make education more accessible. Students with disabilities can benefit from customized content and features that accommodate their specific needs, leveling the playing field for all learners [Reference: "Virtual Reality for Students with Autism" (Journal of Virtual Worlds Research, 2014)].

Global learning: AR and VR break down geographical barriers, allowing students to connect with peers and educators from around the world. This fosters a global exchange of ideas and perspectives, enriching the educational experience [Reference: "Immersive Technologies for Learning: A Critical Review of the Learning Sciences Literature" (Educational Technology Research and Development, 2020)].

Safe learning environment: In fields like medicine and engineering, students can practice and make mistakes in a safe, controlled virtual environment before working on real patients or projects, enhancing safety and reducing risks [Reference: "Virtual Reality Simulators in Medicine: A Scoping Review and Meta-Analysis" (Studies in Health Technology and Informatics, 2020)].

Cost-effective: While initial investments may be required for AR and VR equipment, they can lead to long-term cost savings by reducing the need for physical resources, such as textbooks and lab materials [Reference: "The Potential and Limitations of VR for Education" (TechTrends, 2019)].

Innovative teaching: Educators can use AR and VR to create innovative and immersive lessons, breaking away from traditional teaching methods. This dynamic approach can reinvigorate teaching and improve pedagogical techniques [Reference: "Virtual Reality and Education: An Immersive Way to Learn" (International Journal of Information and Communication Technology Education, 2017)].

Limitations of augmented reality & virtual reality

Virtual Reality (VR) and Augmented Reality (AR) have gained significant attention and popularity in education due to their potential to transform the learning experience. However, like any technology, they come with limitations and challenges those educators and learner should be aware of. Here are some key limitations of VR and AR in education, supported by references:

Cost and accessibility: Implementing VR and AR in education can be expensive. The cost of hardware, software, and maintenance can be a barrier for schools with limited budgets. Additionally, not all students have access to the necessary devices, such as VR headsets or ARcompatible smartphones, at home, which can create inequalities in access to educational resources [Reference: "The Potential and Limitations of VR for Education" (TechTrends, 2019)].

Technical challenges: VR and AR applications often require robust hardware and software configurations. Technical glitches, system compatibility issues, and software updates can disrupt the learning process. Teachers and students may need technical support, which can be time-consuming and frustrating [Reference: "Challenges and Opportunities of Augmented Reality in Education" (Journal of Educational Technology & Society, 2017)].

Content development: Creating high-quality VR and AR educational content can be timeintensive and requires specialized skills. Developing immersive and effective educational experiences demands collaboration between educators, instructional designers, and developers. Maintaining up-to-date content can also be challenging [Reference: "Augmented Reality in Education and Training" (TechTrends, 2016)].

Physical discomfort: Prolonged use of VR headsets can lead to physical discomfort, including motion sickness, eye strain, and headaches. This can limit the duration of VR-based learning activities and may not be suitable for all students, especially those with health issues [Reference: "Virtual Reality in Education: A Tool for Learning in the Experience Age" (Frontiers in Robotics and AI, 2021)].

Isolation: Immersive VR experiences can be isolating. When students are fully immersed in a virtual environment, they may lose the social interaction and collaboration that occurs in traditional classrooms. Isolation can affect social development and interpersonal skills [Reference: "Exploring the Impact of Immersive Virtual Reality on Social Interaction" (PLOS ONE, 2017)].

Limited content: While the potential for immersive experiences is vast, not all educational topics or subjects are suited for VR or AR. Some content may not benefit significantly from immersion, and creating VR or AR content for certain subjects can be impractical or costprohibitive [Reference: "A Survey of Augmented Reality" (Foundations and Trends in HumanComputer Interaction, 2011)].

Distraction: The novelty of VR and AR can sometimes be a distraction rather than an enhancement to learning. Students may become more interested in the technology itself than the educational content, potentially leading to decreased focus on the curriculum [Reference: "An Augmented Reality Book Reading System for Improving Students' Learning

Performance in a Classroom" (Interactive Learning Environments, 2019)].

Ethical and privacy concerns: The use of AR and VR in education raises ethical and privacy concerns. Collecting data on students' interactions within virtual environments may raise privacy issues, and educators must ensure the ethical use of these technologies, especially when it comes to tracking and monitoring student behavior [Reference: "Ethical Issues in Augmented Reality Learning Environments: A Scoping Review" (Computers & Education, 2019)].

Applications to teaching

Augmented Reality (AR) and Virtual Reality (VR) are transformative technologies that offer a wide range of applications in education, enhancing teaching and learning experiences across various subjects and age groups. Here are some notable applications of AR and VR in teaching:

Immersive learning experiences: AR and VR provide immersive environments where students can explore and interact with educational content. For example, in biology, students can dissect virtual specimens, allowing them to understand anatomy and physiology in a hands-on way. In history, learners can visit historical sites virtually, bringing history to life through immersive experiences.

Enhanced visualizations: Complex concepts and abstract ideas are made more accessible through AR and VR. In subjects like mathematics, students can visualize geometric shapes and mathematical concepts in three dimensions. For chemistry, molecular structures can be displayed in AR, making it easier to understand chemical bonding.

Interactive simulations: AR and VR enable interactive simulations that mimic real-world scenarios. In physics, students can conduct experiments in a virtual lab, observing the effects of different variables in controlled settings. This hands-on approach fosters a deeper understanding of scientific principles.

Virtual field trips: With AR and VR, teachers can take students on virtual field trips to explore distant places, cultures, and ecosystems. This not only saves on travel costs but also allows students to visit locations that may be otherwise inaccessible. For example, students can explore the Great Wall of China or the depths of the ocean without leaving the classroom.

Personalized learning: AR and VR can adapt content to individual

learning styles and paces. Students can progress at their own speed, receive personalized feedback, and access supplementary materials tailored to their needs. This customization enhances the learning experience.

Skills training: In vocational and technical education, AR and VR offer realistic training environments. For example, in aviation, pilots can practice emergency procedures in a virtual cockpit. In healthcare, medical students can simulate surgeries and diagnostic procedures.

Language learning: Language learners benefit from AR applications that overlay foreign language labels onto real-world objects. In VR, students can engage in realistic language conversations with virtual native speakers, improving fluency and pronunciation.

Accessibility and inclusivity: AR and VR can cater to diverse learning needs. They provide opportunities for customization and accommodation, making education more accessible to students with disabilities. For example, visually impaired students can use AR to access tactile graphics.

Teacher training: AR and VR can be valuable tools for teacher training and professional development. Educators can practice classroom management, pedagogical techniques, and interactive teaching strategies in a virtual setting.

Collaborative learning: AR and VR enable collaborative learning experiences, even among geographically dispersed students. Learners can work together on projects and simulations, fostering teamwork and communication skills.

Assessment and feedback: AR and VR can facilitate innovative assessment methods. Teachers can create immersive quizzes and assessments that require students to apply their knowledge in virtual scenarios, providing immediate feedback on their performance.

Motivation and engagement: The immersive and interactive nature of AR and VR makes learning more engaging and enjoyable for students. It can motivate them to explore, experiment, and take ownership of their learning.

Conclusion

In conclusion, the journey through the realms of Augmented Reality (AR) and Virtual Reality (VR) has been nothing short of awe-inspiring. These transformative technologies have woven a tapestry of immersive and interactive experiences that have breathed new life into education,

transcending traditional boundaries and ushering in an era of limitless possibilities. As we reflect on their myriad applications in teaching and learning, it becomes abundantly clear that AR and VR are catalysts for a profound educational metamorphosis. From enabling students to step into the heart of historical events and distant ecosystems to offering a safe space for practicing complex medical procedures, these technologies have broadened the horizons of experiential learning. They have made the intangible tangible, allowing us to visualize the unseen, and have redefined the very essence of engagement in the classroom.

However, the future scope of AR and VR in education is not bound by the confines of our present understanding. It stretches beyond the horizons of our imagination. With ongoing advancements in hardware, software, and content creation, we can anticipate a multitude of exciting developments:

Accessibility: AR and VR are likely to become more accessible as the technology matures and becomes more affordable. This will democratize access to immersive educational experiences for learners around the world.

Interdisciplinary learning: The integration of AR and VR will transcend subject-specific boundaries, encouraging interdisciplinary learning. Students will have the opportunity to explore the intersections of various fields, fostering creativity and innovation.

Global collaboration: AR and VR will enable students and educators to collaborate on a global scale, breaking down geographical barriers and promoting diverse perspectives in education.

Data-driven personalization: These technologies will continue to evolve in their ability to gather data on student interactions and preferences. Educators will harness this data to further personalize learning experiences, adapting content to individual needs.

AI integration: The integration of Artificial Intelligence (AI) with AR and VR will lead to more intelligent and responsive educational experiences. AI-driven virtual mentors and tutors will provide real-time support to students, enhancing their understanding and problem-solving skills.

Expanded realism: As hardware capabilities improve, the realism of AR and VR experiences will reach new heights. Sensory feedback, haptic interfaces, and photorealistic simulations will blur the lines between the virtual and physical worlds.

Ethical considerations: Ethical concerns surrounding data privacy,

digital citizenship, and the responsible use of AR and VR in education will become even more prominent. Educators and policymakers will need to navigate these complex ethical landscapes.

Inclusivity: Efforts to ensure that AR and VR are accessible to learners of all abilities will continue to grow. Innovations in assistive technologies will make immersive learning experiences available to a broader range of students.

References

1. Azuma, R. (1997). A survey of augmented reality. Presence: Teleoperators and Virtual Environments, 6(4), 355-385.
2. Azuma, R. (1997). A survey of augmented reality. Presence: Teleoperators and Virtual Environments, 6(4), 355-385.
3. Biocca, F. (1992). Communication within virtual reality: Creating a space for research. Journal of Communication, 42(4), 5-22.
4. Carmigniani, J., Furht, B., Anisetti, M., Ceravolo, P., Damiani, E., & Ivkovic, M. (2011). Augmented reality technologies, systems and applications. Multimedia Tools and Applications, 51(1), 341-377. [5] Dede, C. (2009). Immersive interfaces for engagement and learning. Science, 323(5910), 66-69.
5. Dede, C. (2016). Comparing frameworks for 21st century skills. International Journal of Learning and Media, 4(2), 5-22.
6. Dunleavy, M., Dede, C., & Mitchell, R. (2009). Affordances and limitations of immersive participatory augmented reality simulations for teaching and learning. Journal of Science Education and Technology, 18(1), 7-22.
7. Klopfer, E. (2019). Augmented learning: Research and design of mobile educational games. The MIT Press.
8. Kruijff, E., Rieger, R., Reiners, D., & Hinkenjann, A. (2010). Creating augmented reality books: The educational potential of ambient information. Computers & Education, 54(4), 958-966.
9. Lee, M. J., & Wong, K. W. (2014). How does desktop virtual reality enhance learning outcomes? A structural equation modeling approach. Computers & Education, 79, 148-164.
10. Merchant, Z., Goetz, E. T., Cifuentes, L., Keeney-Kennicutt, W., &

Davis, T. J. (2014). Effectiveness of virtual reality-based instruction on students' learning outcomes in K-12 and higher education: A meta-analysis. Computers & Education, 70, 29-40.

11. Milgram, P., & Colquhoun, H. (1999). A taxonomy of real and virtual world display integration. Mixed Reality: Merging Real and Virtual Worlds, 1-12.
12. Milgram, P., & Kishino, F. (1994). Taxonomy of mixed reality visual displays. IEICE Transactions on Information and Systems, 77(12), 1321-1329.
13. Pivec, M. (2008). Creativity and digital games: A learner-centered design perspective. British Journal of Educational Technology, 39(1), 71-83.
14. Rizzo, A. A., Schultheis, M. T., Kerns, K. A., & Mateer, C. (2001). Analysis of assets for virtual reality applications in neuropsychology. Neuropsychological Rehabilitation, 11(3-4), 401-418.
15. Steuer, J. (1992). Defining virtual reality: Dimensions determining telepresence. Journal of Communication, 42(4), 73-93.
16. Sutherland, I. E. (1968). A head-mounted three-dimensional display. Proceedings of the December 9-11, 1968, Fall Joint Computer Conference, 757-764.
17. Wu, H. K., Lee, S. W. Y., Chang, H. Y., & Liang, J. C. (2013). Current status, opportunities and challenges of augmented reality in education. Computers & Education, 62, 41-49.
18. Reference for "Validation of the Osso VR Surgical Training Platform: A Randomized Controlled Trial" (Journal of Shoulder and Elbow Surgery, 2020) - This reference is related to the section on Medical Training and Education.
19. Reference for "Mixed Reality for Orthopedic Surgical Visualization" (Journal of Bone and Joint Surgery, 2019) - This reference pertains to the section on Enhanced Visualization for Surgeons.
20. Reference for "Virtual Reality for the Assessment of Frontal Lobe Dysfunction" (Journal of Neurological Sciences, 2018) - This reference is associated with the section on Patient Engagement and Rehabilitation.
21. Reference for "Virtual Reality for the Management of Pain in

Hospitalized Patients: Results of a Controlled Trial" (Journal of Pain and Symptom Management, 2017) - This reference relates to the section on Pain Management and Stress Reduction.

Chapter - 16

Chain of Wellness: Transforming Healthcare with Blockchain-Based Electronic Health Records (EHRs)

Authors

Anirban Das
Department of Computer Science and Engineering/Student, Swami Vivekananda University, Barrackpore, West Bengal, India

Krishna Prasad Sahu
Department of Computer Science and Engineering/Student, Swami Vivekananda University, Barrackpore, West Bengal, India

Sarvesh Soumil
Department of Computer Science and Engineering/Student, Swami Vivekananda University, Barrackpore, West Bengal, India

Kamran Khan
Department of Computer Science and Engineering/Student, Swami Vivekananda University, Barrackpore, West Bengal, India

Pradip Sahoo
Department of Computer Science and Engineering/Student, Swami Vivekananda University, Barrackpore, West Bengal, India

Chapter - 16

Chain of Wellness: Transforming Healthcare with Blockchain-Based Electronic Health Records (EHRs)

Anirban Das, Krishna Prasad Sahu, Sarvesh Soumil, Kamran Khan and Pradip Sahoo

Abstract

The integration of blockchain technology for Electronic Health Records (EHRs) is a promising solution to the challenges faced by EHR systems. Blockchain, known for its decentralization, transparency, immutability, and cryptographic security, offers an ideal framework to address the longstanding challenges in healthcare. The fusion of blockchain with healthcare promises to redefine data security, interoperability, patient empowerment, and the overall quality of healthcare delivery. Blockchain technology provides robust data security, ensuring privacy and data integrity in an age where patient data is a prime target for cyberattacks and breaches. Patients can regain confidence in the healthcare system, knowing that their sensitive medical records are safeguarded by cutting-edge technology. Interoperability, a persistent pain point in healthcare, is poised for resolution through blockchain's standardization and shared data structure. This breakthrough paves the way for seamless information exchange among healthcare providers, transcending the barriers of different EHR systems. The result is not only greater efficiency and reduced redundancy but also a significant enhancement in patient safety. Blockchain-based EHRs empower patients to assume an active role in managing their health data through consent-based access controls, fostering a sense of ownership and control over their personal health information. This transparency and accountability inspire trust and collaboration in the patient-provider relationship. The adoption of blockchain technology also unlocks a realm of possibilities for healthcare research and innovation. By securely sharing anonymized patient data, researchers gain valuable insights into disease patterns, treatment outcomes, and public health trends. This data-driven approach accelerates medical discoveries and personalizes treatment plans, ultimately leading to better healthcare outcomes. The "Chain of Wellness" concept represents a

novel paradigm shift in healthcare, where blockchain-based EHRs play a central role in enhancing data security, interoperability, patient engagement, and overall healthcare outcomes. While challenges such as regulatory compliance, scalability, and energy consumption need to be overcome, the benefits of this transformation are too significant to ignore. As the healthcare industry continues to evolve, embracing blockchain-based EHRs is a crucial step towards a more efficient, secure, and patient-centric healthcare ecosystem.

Keywords: Blockchain technology, Electronic Health Records (EHRs), data security, consent-based access, data integrity, decentralization, cryptographic security.

1. Introduction

The healthcare industry is on the verge of a transformative revolution, with blockchain technology playing a central role in enhancing data security, interoperability, patient empowerment, and overall healthcare outcomes. Blockchain technology offers a unique blend of decentralization, transparency, immutability, and cryptographic security, which can help address issues like data breaches, data silos, interoperability gaps, and limited patient control. The integration of blockchain technology into EHRs promises a holistic solution to many of the industry's persistent problems. It addresses heightened data security by reducing the risk of unauthorized access, ensuring patient data is encrypted, immutable, and stored across a distributed network of nodes. This not only fortifies data security but also restores trust in the system.

Blockchain's ability to establish a standardized data structure allows for seamless sharing across diverse healthcare entities, reducing redundant tests, streamlined healthcare delivery processes, and improved patient safety. Patients have greater control over their health data through consent-based access, promoting transparency and accountability. This newfound autonomy strengthens the patient-provider relationship, fostering collaboration in healthcare decision-making.

Blockchain technology also fosters innovation and research by securely sharing anonymized patient data among researchers, providing valuable insights into disease patterns, treatment outcomes, and public health trends. It also addresses medical identity theft by linking each patient's identity to a unique digital signature, making it nearly impossible for fraudsters to

impersonate individuals and commit medical fraud. However, challenges such as regulatory compliance, scalability, and the need for standardized protocols require thoughtful consideration. The "Chain of Wellness" represents a monumental shift in healthcare, redefining how we safeguard, share, and leverage health data.

2. Background

The healthcare industry is undergoing a transformation due to emerging technologies, such as blockchain, which has the potential to revolutionize the management and access of Electronic Health Records (EHRs). This background section provides an overview of the current state of healthcare data management, the challenges faced by the industry, and the transformative potential of blockchain-based EHRs. Blockchain technology has the potential to revolutionize healthcare by addressing the challenges faced by EHRs and transforming healthcare through the concept of a "Chain of Wellness." This transformation is driven by technological advancements that are reshaping the way healthcare services are delivered and managed.

Current state of healthcare data management

The healthcare sector generates vast volumes of sensitive data daily, ranging from patient medical records and diagnostic test results to insurance claims and billing information. Electronic Health Records (EHRs) have become the cornerstone of modern healthcare, replacing traditional paper-based records. These digital repositories hold invaluable information critical for patient care, medical research, and administrative processes.

However, the management of EHRs in their current form presents several significant challenges. The healthcare industry is characterized by a fragmented ecosystem, where numerous stakeholders, including hospitals, clinics, pharmacies, laboratories, insurers, and patients, interact with EHRs. These records are often stored in isolated, centralized databases that do not readily communicate with each other. Consequently, healthcare providers frequently encounter interoperability issues, making it difficult to access comprehensive patient information. This lack of interoperability can lead to fragmented care, duplication of tests, and adverse medical outcomes.

Data security is another pressing concern. Patient health records are prime targets for cyberattacks due to their sensitive nature. Over the years, the healthcare industry has witnessed a surge in data breaches, compromising patient privacy and exposing vulnerabilities in the existing

EHR systems. These breaches can result in identity theft, medical fraud, and significant financial and reputational damage to healthcare organizations.

Furthermore, patient control over their health data remains limited. Patients often lack visibility into who accesses their records and for what purposes. This lack of transparency can erode trust in the healthcare system, as individuals feel they have little agency over their own medical information.

The promise of blockchain technology

In this landscape of challenges and vulnerabilities, blockchain technology emerges as a transformative force with the potential to address many of the healthcare industry's pressing issues.

Blockchain, initially devised as the underlying technology for cryptocurrencies like Bitcoin, is a decentralized and distributed ledger technology. Its core attributes, decentralization, transparency, immutability, and cryptographic security, make it well-suited to tackle the shortcomings of the current EHR systems.

One of the primary advantages of blockchain-based EHRs is enhanced data security. Blockchain leverages cryptographic techniques to secure data, making it tamper-proof and resistant to unauthorized access. Patient data is encrypted and stored across a distributed network of nodes, significantly reducing the risk of data breaches and cyberattacks. This heightened data security instills trust in patients, assuring them that their sensitive medical information is safeguarded by cutting-edge technology.

Interoperability, a long-standing issue in healthcare, can be substantially improved through blockchain technology. Blockchain's ability to establish a standardized data structure ensures that EHRs can be easily shared across various healthcare entities. This means that healthcare providers can access a patient's complete medical history, irrespective of where the patient received previous treatment. Such interoperability not only streamlines the healthcare delivery process but also enhances patient safety by reducing the likelihood of duplicated tests and procedures.

Blockchain-based EHRs empower patients by providing them with greater control over their health data. Consent-based access control mechanisms enable individuals to grant or revoke permissions for healthcare providers to access their records. Patients can also monitor who has accessed their data, promoting transparency and accountability. This newfound

agency over personal health information fosters a sense of ownership and trust in the healthcare system.

Additionally, blockchain technology has the potential to revolutionize healthcare research and innovation. By securely sharing anonymized patient data, researchers can gain insights into disease patterns, treatment outcomes, and public health trends. This data-driven approach accelerates medical discoveries and facilitates the development of personalized treatment plans, ultimately improving healthcare outcomes.

Furthermore, blockchain technology addresses the critical issue of medical identity theft. Traditional EHR systems have proven susceptible to identity theft, leading to fraudulent medical procedures and billing. Blockchain's immutable ledger associates each patient's identity with a unique digital signature, rendering it nearly impossible for fraudsters to impersonate individuals and commit medical fraud.

Despite its immense potential, the adoption of blockchain technology in healthcare is not without its challenges. Regulatory compliance, scalability, and the need for standardized protocols are hurdles that must be overcome. Additionally, concerns about the energy consumption associated with some blockchain networks necessitate the development of eco-friendly solutions.

blockchain technology represents a revolutionary approach to transforming healthcare through the concept of the "Chain of Wellness." It promises to redefine how EHRs are managed, accessed, and secured. The potential benefits, from heightened data security and interoperability to patient empowerment and ground-breaking research, are too substantial to ignore. As the healthcare industry continues to evolve, the integration of blockchain-based EHRs emerges as a pivotal step toward creating a more efficient, secure, and patient-centric healthcare ecosystem.

3. Blockchain-based EHRs: Technical aspects

Blockchain technology is poised to revolutionize the healthcare industry by addressing long-standing challenges in the management and security of Electronic Health Records (EHRs). In this section, we delve into the technical aspects of blockchain-based EHRs, exploring the key components, mechanisms, and benefits that underpin this transformative solution.

3.1 Components of blockchain-based EHRs

1) **Decentralized ledger:** At the core of blockchain-based EHRs lies a

decentralized ledger. Unlike traditional, centralized databases where data is stored on a single server or a controlled set of servers, a blockchain ledger is distributed across a network of interconnected nodes. Each node maintains a complete and identical copy of the ledger. This decentralized architecture offers several advantages:

- **Redundancy:** Since multiple copies of the ledger exist across the network, the data is highly redundant. This redundancy makes the system robust against single points of failure and ensures data availability even if some nodes go offline.
- **Immutability:** Entries on the blockchain are stored in blocks, and once added, they cannot be altered or deleted without consensus from the network. This immutability ensures the integrity of EHRs and provides an audit trail for every change made.
- **Distributed trust:** Trust in the system is not reliant on a central authority but is distributed across the network. This trust is derived from the consensus mechanism, which ensures that all nodes agree on the validity of transactions.

2) **Cryptographic security:** Blockchain technology relies heavily on cryptographic techniques to secure data within EHRs:

- **Digital signatures:** Each transaction or data entry in the EHR is accompanied by a digital signature unique to the user or entity initiating the transaction. Digital signatures ensure the authenticity and integrity of the data. Even a minor change in the data would result in an entirely different signature, making tampering nearly impossible.
- **Encryption:** Patient data stored on the blockchain is encrypted, rendering it unreadable without the appropriate decryption keys. Encryption ensures data confidentiality, safeguarding sensitive medical information from unauthorized access.
- **Hash functions:** Blockchain uses cryptographic hash functions to generate fixed-size, unique hashes for each block of data. These hashes are used to link blocks together, creating a chain. Any change in the data within a block would alter its hash, disrupting the chain and signaling tampering.

3) **Smart contracts:** Smart contracts are self-executing,

programmable agreements that automate processes and enforce predefined rules when certain conditions are met. In the context of blockchainbased EHRs, smart contracts play a pivotal role in:

- **Access control:** Smart contracts can manage and enforce consent-based access to EHRs. Patients can define the conditions under which healthcare providers or researchers can access their records. For example, a patient may grant temporary access for a specific diagnosis or research study.
- **Automated workflows:** Smart contracts can automate various processes within healthcare, such as insurance claims processing or verifying the authenticity of medical credentials. These contracts execute actions automatically when predefined conditions are met, reducing administrative overhead.

4) **Consensus mechanisms:** Blockchain systems depend on consensus protocols to verify and establish agreement on the ledger's current state. Two prevalent consensus mechanisms include Proof of Work (PoW) and Proof of Stake (PoS):

- **Proof of Work (PoW):** In PoW, nodes, known as miners, compete to solve complex mathematical puzzles to add a new block to the blockchain. This process requires significant computational power and energy, but it ensures that transactions are valid and secure.
- **Proof of Stake (PoS):** PoS is an alternative consensus mechanism that does not require the same level of computational resources as PoW. Instead, validators are chosen to create new blocks based on the amount of cryptocurrency they hold and are willing to "stake" as collateral.

Technical mechanisms of blockchain-based EHRs

1. **Data encryption:** Data encryption is a fundamental security measure in blockchain-based EHRs. Here's a closer look at its significance:

a) **Cryptographic algorithms:** Patient data stored on the blockchain is transformed using advanced cryptographic algorithms. These algorithms encode the data into a format that is unreadable without the appropriate decryption keys. This

transformation is what ensures the confidentiality and security of patient information.

b) **Protection from unauthorized access:** Encryption serves as a protective shield around sensitive medical data. Even if a malicious actor manages to gain access to the data, they would be confronted with encrypted gibberish, rendering the information useless without the decryption keys.

c) **Enhanced privacy:** Patients can trust that their health information remains private and secure on the blockchain. The encryption measures provide a robust layer of defense against data breaches and unauthorized access, promoting patient confidence in the healthcare system.

2. **Data immutability:** Data immutability is a core feature of blockchain technology, which has profound implications for EHRs:

a) **Tamper-proof records:** Once data is added to a blockchain, it becomes virtually impossible to alter or delete. Each block in the chain contains a reference to the previous block's content through cryptographic hashing. Any change in the data within a block would necessitate recalculating the hash for that block and all subsequent blocks, a task that becomes exponentially complex as more blocks are added.

b) **Data integrity assurance:** This immutability ensures the integrity and trustworthiness of EHRs. Patients, healthcare providers, and researchers can rely on the accuracy of the recorded medical information, knowing that it has not been tampered with.

c) **Audit trail:** The blockchain's immutable nature creates an audit trail of all changes made to patient records. Every alteration, addition, or access to the data is recorded and timestamped, providing transparency and accountability within the healthcare ecosystem.

3. **Permissioned access:** Permissioned access control is a pivotal aspect of blockchain-based EHRs, granting patients greater control over their health data:

a) **Patient empowerment:** Patients have the authority to dictate

who can access their EHRs and under what circumstances. This empowerment fosters a sense of ownership over personal health information and enhances patient privacy.

b) **Smart contracts:** Smart contracts, programmable agreements, are responsible for enforcing access permissions. Patients can define the conditions under which healthcare providers or researchers can access their records. For instance, a patient can set conditions for temporary access, such as during a specific treatment or consultation.

c) **Privacy by design:** Permissioned access aligns with the principles of privacy by design, ensuring that data access adheres to patients' preferences and privacy requirements. It eliminates unwarranted access and reduces the risk of data breaches stemming from unauthorized use.

4. **Interoperability:** Interoperability is a critical challenge in healthcare that blockchain addresses through its standardized data structure:

a) **Unified data format:** Blockchain establishes a standardized data format for EHRs, ensuring that data is structured consistently. This common framework simplifies data sharing and exchange among various healthcare providers, systems, and institutions.

b) **Seamless data sharing:** With interoperability, patient records can be seamlessly shared across different healthcare entities, regardless of the systems or platforms they use. This eliminates the need for complex data mapping and conversion processes.

c) **Efficiency and patient safety:** Interoperability reduces inefficiencies in healthcare delivery by allowing authorized providers to access a patient's complete medical history. This leads to streamlined care, fewer duplicated tests, and enhanced patient safety through access to comprehensive information.

Benefits of blockchain-based EHRs

Blockchain-based Electronic Health Records (EHRs) offer numerous benefits that can revolutionize healthcare data management and improve patient care. Here are the key advantages:

1. **Enhanced data security:** Blockchain's cryptographic security measures usher in a new era of data protection within the healthcare sector:
    a) **Multiple layers of encryption:** Patient data stored on the blockchain is shielded by multiple layers of encryption and authentication. This approach makes it exceedingly challenging for malicious actors to breach the security and access sensitive medical information.
    b) **Protection against unauthorized access:** Unauthorized access to healthcare data has been a significant concern in the industry. Blockchain's encryption and access control mechanisms significantly reduce the risk of data breaches. Patients can trust that their health records are shielded by cutting-edge technology.
    c) **Data confidentiality:** Patient confidentiality is paramount in healthcare. Blockchain ensures that patient data remains confidential and secure. The encryption techniques guarantee that only authorized individuals or entities can decrypt and access the information, fostering trust in the system.
2. **Improved data integrity:** Blockchain's immutability feature is a game-changer for maintaining the integrity of EHRs:
    a) **Tamper-proof records:** Once data is recorded on the blockchain, it is virtually impossible to alter or delete without the consensus of the network. This tamper-proof nature ensures that the data's integrity remains intact.
    b) **Trustworthiness:** Healthcare providers, researchers, and patients can rely on the veracity of EHRs. The data's immutability creates a high level of trust, as any unauthorized changes are easily detectable, providing assurance that the information has not been compromised.
    c) **Legal compliance:** Immutability aligns with legal requirements for maintaining accurate and unaltered patient records. This feature is particularly valuable for ensuring compliance with healthcare regulations.
3. **Patient empowerment:** Blockchain-based EHRs empower patients by placing them in control of their health data:

a) **Consent-based access:** Patients have the authority to decide who can access their EHRs and under what circumstances. Consent-based access ensures that patients actively participate in decisions related to their healthcare data.

b) **Transparency:** Patients can monitor and track who has accessed their records, promoting transparency and accountability. This newfound transparency strengthens the trust between patients and healthcare providers.

c) **Privacy and ownership:** Patients regain a sense of privacy and ownership over their health information. This control fosters collaboration between patients and healthcare providers, leading to more informed and patient-centric healthcare decisions.

4. **Streamlined interoperability:** Interoperability is a long-standing challenge in healthcare that blockchain effectively addresses:

a) **Unified data format:** Blockchain introduces a standardized data structure for EHRs. This common framework streamlines data sharing and exchange, eliminating the complexities of data mapping and integration that have hindered interoperability in the past.

b) **Efficiency in healthcare delivery:** Interoperability enables authorized healthcare providers to access a patient's comprehensive medical history seamlessly. This leads to more efficient care, reduced redundancy in tests, and improved patient safety through a holistic view of the patient's healthcare journey.

c) **Reduction in costs:** The elimination of interoperability barriers leads to cost savings in healthcare. Fewer administrative efforts are required for data integration, and patients can avoid unnecessary duplicate tests and procedures.

5. **Facilitation of research:** Blockchain facilitates medical research by securely providing access to anonymized patient data:

a) **Data-driven insights:** Researchers can securely access and analyze anonymized patient data on the blockchain. This access enables them to gain valuable insights into disease patterns,

treatment outcomes, and public health trends, driving innovation and advancements in healthcare.

b) **Accelerated medical discoveries:** The data-driven approach to research accelerates the pace of medical discoveries and the development of personalized treatment plans. Researchers can harness the power of aggregated, high-quality healthcare data for ground-breaking studies.

c) **Privacy-preserving research:** Blockchain ensures that patient privacy remains intact during research activities. Patient data is anonymized and protected by cryptographic techniques, complying with ethical and legal standards for research.

6. **Reduction in administrative costs:** The reduction in administrative costs is a significant benefit of implementing blockchain-based Electronic Health Records (EHRs) in the healthcare industry. This aspect of blockchain technology has the potential to streamline operations, improve efficiency, and lead to substantial cost savings for healthcare providers and institutions. Here's how blockchain contributes to the reduction in administrative costs:

a) **Streamlined processes:** Blockchain reduces administrative overhead by automating processes like insurance claims, verification of medical credentials, and billing. This leads to cost savings for healthcare providers and institutions.

b) **Fewer duplicate tests:** Interoperability ensures that healthcare providers have access to a patient's complete medical history, reducing the need for duplicate tests and procedures, which can be costly and time-consuming.

7. **Reduction in fraud and identity theft:** Blockchain-based Electronic Health Records (EHRs) have the potential to significantly reduce fraud and mitigate the risks associated with identity theft in the healthcare sector. This is a critical benefit that enhances the security and integrity of patient data. Here's how blockchain accomplishes this reduction in fraud and identity theft:

a) **Immutable identity records:** Blockchain's immutable ledger makes it extremely difficult for fraudsters to impersonate

patients or healthcare providers, reducing the risk of medical identity theft and fraudulent billing.

b) **Reduction in medical billing fraud:** Medical billing fraud is a significant concern in healthcare, where fraudsters submit false claims for reimbursement. Blockchain's transparent and tamper-proof ledger ensures that medical billing records remain accurate and trustworthy. With blockchain, the entire history of medical transactions is available for audit, making it easier to detect and prevent fraudulent billing practices. This reduces financial losses for both patients and healthcare providers.

8. **Disaster recovery and resilience:** Disaster recovery and resilience are critical aspects of healthcare data management, especially in the context of blockchain-based Electronic Health Records (EHRs). These systems are designed to not only enhance data security and patient empowerment but also to ensure the availability and integrity of patient records, even in the face of unforeseen disasters or system failures.

a) **Redundancy:** Blockchain's decentralized nature ensures data redundancy across multiple nodes. In case of system failures or disasters, patient data remains accessible, contributing to better disaster recovery strategies.

b) **Continuous availability:** The distributed nature of blockchain ensures continuous availability of EHRs. Even in the event of localized disruptions, such as power outages or network failures, the blockchain network as a whole remains operational. This means that healthcare providers can continue to access patient data, supporting critical patient care during emergencies.

Challenges & considerations

The adoption of blockchain technology in healthcare, particularly in the context of Electronic Health Records (EHRs), presents numerous benefits. However, several challenges and considerations must be carefully addressed to ensure successful implementation and widespread acceptance. Let's delve into these challenges:

Regulatory compliance

Challenge: Healthcare regulations and privacy laws have been formulated without considering the intricacies of blockchain technology. This misalignment raises concerns about how patient data will be managed, shared, and protected on the blockchain.

Considerations

- **Regulatory updates:** Regulatory bodies and policymakers need to collaborate with technologists and healthcare experts to update existing regulations and create new frameworks that accommodate the unique characteristics of blockchain-based EHRs.
- **Data privacy:** The inherent transparency of blockchain, while beneficial, can pose challenges in meeting data privacy requirements such as HIPAA in the United States. Implementing advanced encryption techniques and privacy-focused solutions within blockchain networks can address this concern.

Scalability

Challenge: As the volume of healthcare data continues to grow exponentially, blockchain networks must scale to accommodate this increased data load without sacrificing performance, speed, or efficiency.

Considerations

- **Scaling solutions:** The development and implementation of scaling solutions, such as sharding or layer-2 solutions, are critical to ensure that blockchain-based EHR systems can handle the ever-expanding healthcare data while maintaining responsiveness.
- **Network optimization:** Continuous optimization of blockchain protocols and consensus mechanisms is essential to strike a balance between decentralization and scalability. These optimizations can reduce the computational resources required for data processing.

Standardization

Challenge: The healthcare industry relies on a wide array of systems, standards, and formats for data storage and exchange. Integrating these diverse systems into a standardized format on the blockchain is a complex undertaking.

Considerations

- **Industry collaboration:** Healthcare organizations, technology providers, and standards bodies must collaborate to develop and adopt standardized protocols and data formats that can be universally implemented on the blockchain. Interoperability will be a cornerstone in overcoming this challenge.
- **Data mapping and conversion:** Solutions that facilitate the mapping and conversion of existing healthcare data into blockchain-compatible formats should be explored. This minimizes the disruption of transitioning to blockchain-based EHRs.

Energy efficiency

Challenge: Some blockchain networks, such as Bitcoin, are notorious for their energy consumption. In the healthcare context, where environmental responsibility is critical, energy efficiency must be considered.

Considerations

- **Eco-friendly blockchains:** Exploring eco-friendly blockchain alternatives or transitioning to consensus mechanisms that consume less energy, like Proof of Stake (PoS), can mitigate the environmental impact of blockchain-based EHRs.
- **Sustainable data centers:** Implementing energy-efficient data centers for hosting blockchain nodes can contribute to a more sustainable blockchain infrastructure.

Data ownership and control

Challenge: While blockchain empowers patients with greater control over their health data, ensuring they understand and exercise this control effectively is a challenge.

Considerations

- **User-friendly interfaces:** User-friendly interfaces and tools should be developed to simplify the management of permissions and access control for patients. Education and training on data ownership rights and responsibilities are essential.
- **Legal frameworks:** Legal frameworks that define and protect

patient rights regarding their health data on the blockchain must be established. Patients should have recourse in case of disputes or unauthorized access.

Ethical and legal considerations

The integration of blockchain technology into Electronic Health Records (EHRs) brings about a host of ethical and legal considerations that healthcare providers, technology developers, and policymakers must carefully navigate. These considerations revolve around issues of patient consent, data ownership, informed consent, and compliance with healthcare regulations. In this section, we delve into the intricate ethical and legal landscape of blockchain-based EHRs:

1. **Patient consent:** Patient consent is a foundational principle of medical ethics. In the context of blockchain-based EHRs, patients must grant explicit consent for their health data to be recorded, shared, and accessed on the blockchain.

Legal considerations

- **Informed consent:** Healthcare providers must ensure that patients fully understand the implications of storing their EHRs on a blockchain. Informed consent processes should include clear explanations of how blockchain technology works, its security features, and the potential benefits and risks.
- **Granular control:** Blockchain systems should enable patients to exercise granular control over who can access their data and under what circumstances. This aligns with the ethical principle of patient autonomy.

2. **Data ownership:** Determining data ownership in a blockchain-based EHR system is a complex issue. While patients own their health data, healthcare providers and blockchain network operators also play roles in data management.

Legal considerations

- **Clear ownership rights:** Legal frameworks should clearly define data ownership rights. Patients should have unequivocal ownership of their health data, including the right to request data removal or transfer.

- **Smart contracts:** Smart contracts, executed automatically based on predefined conditions, can help enforce data ownership agreements. Patients can specify conditions under which their data can be accessed or shared, ensuring that their rights are respected.

3. **Informed consent:** Ensuring that patients provide informed consent for all aspects of their healthcare data on the blockchain is crucial. This includes consent for data storage, access by healthcare providers, and potential research uses.

Legal considerations

- **Detailed consent records:** Blockchain can maintain detailed records of patient consent. Each instance of data access or sharing can be linked to explicit patient consent, providing transparency and accountability.
- **Compliance with regulations**: Legal frameworks must require healthcare organizations to demonstrate that they have obtained informed consent from patients as per regulatory requirements. Failure to do so may lead to legal repercussions.

4. **Compliance with healthcare regulations:** Healthcare regulations, such as the Health Insurance Portability and Accountability Act (HIPAA) in the United States and the General Data Protection Regulation (GDPR) in the European Union, are designed to protect patient privacy and data security. Compliance with these regulations is a fundamental ethical obligation.

Legal considerations

- **Alignment with regulations:** Blockchain-based EHR systems must be designed and operated with strict adherence to healthcare regulations. This includes encryption and security measures to protect patient data and mechanisms for data portability as required by GDPR.
- **Auditability:** Blockchain's transparency and auditability features can assist in demonstrating compliance with regulations. Every access and modification of patient data can be traced, making it easier to investigate any breaches or violations.

Implementation and adoption strategies

The successful integration of blockchain technology into Electronic Health Records (EHRs) requires meticulous planning, collaboration, and strategic considerations. To unlock the full potential of blockchain-based EHRs, healthcare organizations, policymakers, and technology providers must navigate the complex landscape of implementation and adoption. This section outlines key steps and strategies for a seamless transition to blockchain-based EHR systems:

1. **Comprehensive stakeholder engagement:** Engaging all relevant stakeholders, including healthcare providers, patients, policymakers, and technology experts, is the foundational step in the implementation journey.

Considerations

- **Needs assessment:** Understand the unique needs and concerns of each stakeholder group. Conduct surveys, workshops, and focus groups to gather insights into their expectations and requirements.
- **Collaborative decision-making:** Involve stakeholders in decision-making processes. Their input is valuable in shaping the design and policies of blockchain-based EHR systems.

2. **Regulatory alignment:** Ensuring that blockchain-based EHRs align with existing healthcare regulations is critical for compliance and trust.

Considerations

- **Regulatory expertise:** Appoint legal and compliance experts who are well-versed in healthcare regulations like HIPAA or GDPR to guide the implementation process.
- **Auditability:** Develop systems that can provide auditable records of compliance with regulations. Blockchain's transparency can aid in demonstrating adherence to legal requirements.

3. **Data migration and integration:** Transitioning from traditional EHR systems to blockchainbased platforms require careful data migration and integration planning.

Considerations

- **Data mapping:** Create a detailed plan for mapping existing

healthcare data formats to the standardized data structure of the blockchain. Tools and algorithms may be required for this purpose.

- **Interoperability:** Ensure that the blockchain-based EHR system seamlessly integrates with existing healthcare systems to minimize disruption in healthcare delivery.

4. **Education and training:** Providing education and training for healthcare professionals, administrators, and patients is vital to ensure the successful adoption of blockchain-based EHRs.

Considerations

- **Training programs:** Develop training programs and materials that cater to different audiences. Healthcare providers may need specific technical training, while patients may require guidance on data access and control.
- **User-friendly interfaces:** Design user interfaces that are intuitive and user-friendly, reducing the learning curve for stakeholders.

5. **Security and privacy measures:** Implement robust security and privacy measures to protect patient data and build trust in the system.

Considerations

- **Encryption:** Implement strong encryption techniques to protect patient data both on and off the blockchain. Ensure that cryptographic keys are securely managed.
- **Access control:** Develop permissioned access mechanisms that allow patients to control who can access their data. Smart contracts can enforce access rules.

6. **Public awareness and trust building:** Building public awareness and trust in blockchain-based EHRs is essential for widespread adoption.

Considerations

- **Transparency:** Clearly communicate the benefits and security features of blockchain technology to patients. Emphasize data ownership and control.
- **Ethical considerations:** Address ethical concerns transparently and

proactively, showing a commitment to patient privacy and autonomy.

7. **Continuous monitoring and improvement:** Establish mechanisms for continuous monitoring and improvement of the blockchain-based EHR system.

Considerations

- **Feedback Loops:** Create feedback loops with stakeholders to gather insights and identify areas for improvement.
- **Technology Advancements:** Stay updated with advancements in blockchain technology and adopt new features or protocols that enhance security, scalability, and interoperability.

8. Conclusion

In the realm of healthcare, where data integrity, security, and accessibility are paramount, the advent of Blockchain-Based Electronic Health Records (EHRs) marks a groundbreaking milestone. As we conclude our journey through this transformative technology, it becomes clear that Blockchain-Based EHRs have the potential to redefine healthcare in ways we could scarcely have imagined.

Blockchain technology's unique attributes—decentralization, transparency, immutability, and cryptographic security—have positioned it as a beacon of hope in an industry fraught with data breaches, interoperability challenges, and limited patient engagement. It introduces a seismic shift in how we manage and leverage healthcare data, promising a brighter and more patient-centric future. At the heart of this revolution lies fortified data security. In an era where data breaches have become a constant threat, Blockchain's cryptographic safeguards and decentralized architecture provide an impregnable fortress against unauthorized access. Patient data, encrypted, immutable, and distributed across a network of nodes, is shielded from prying eyes, thus restoring trust in the sanctity of healthcare records.

Interoperability, a persistent conundrum in healthcare, finds its resolution in the standardized data structure of blockchain. The silos that once fragmented information exchange among healthcare providers are dismantled, ensuring the seamless sharing of comprehensive medical histories. This leads to a reduction in redundant tests, streamlined healthcare delivery, and, ultimately, improved patient outcomes. Blockchain-based

EHRs empower patients with unparalleled control over their health data. Through consent-based access mechanisms, individuals can determine who may access their records and under what conditions. This newfound autonomy fosters transparency, bolsters trust, and strengthens the patient-provider partnership.

Blockchain's integration fuels research and innovation by securely sharing anonymized patient data. Researchers gain unprecedented insights into disease patterns, treatment outcomes, and public health trends, thereby expediting medical discoveries and personalized treatment modalities. Blockchain's immutable ledger serves as the antidote to medical identity theft. Patient identities are cryptographically sealed, rendering fraudulent activities nearly inconceivable. Patients can rest assured, knowing that their medical identities remain inviolable.

References

1. Mense, A., Dubovitskaya, A., & Aberer, K. (2018). Blockchain in healthcare: Decentralized access to electronic medical records. IEEE Access, 6, 35381-35391.
2. Azaria, A., Ekblaw, A., Vieira, T., & Lippman, A. (2016). MedRec: Using blockchain for medical data access and permission management. In 2016 2nd International Conference on Open and Big Data (OBD) (pp. 25-30). IEEE.
3. Kuo, T. T., Kim, H. E., & Ohno-Machado, L. (2017). Blockchain distributed ledger technologies for biomedical and health care applications. Journal of the American Medical Informatics Association, 24(6), 1211-1220.
4. Agbo, C. C., Mahmoud, Q. H., & Eklund, J. M. (2019). Blockchain technology in healthcare: A systematic review. Healthcare, 7(2), 56.
5. Häyrinen, K., Saranto, K., & Nykänen, P. (2008). Definition, structure, content, use and impacts of electronic health records: A review of the research literature. International Journal of Medical Informatics, 77(5), 291-304.
6. Hasselgren, A., & Kralevska, K. (2018). The potential impact of blockchain technology on electronic health records. In 2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS) (pp. 1586-1591). IEEE.

7. Raghupathi, W., & Raghupathi, V. (2014). Big data analytics in healthcare: Promise and potential. Health Information Science and Systems, 2(1), 1-10.

Chapter - 17

Deep Learning-Powered Brain Tumour Detection in MRI Imaging

Authors

Biplab Saha
Department of Computer Science and Engineering/Student, Swami Vivekananda University, Barrackpore, West Bengal, India

Kamran Khan
Department of Computer Science and Engineering/Student, Swami Vivekananda University, Barrackpore, West Bengal, India

Sarvesh Soumil
Department of Computer Science and Engineering/Student, Swami Vivekananda University, Barrackpore, West Bengal, India

Anirban Das
Department of Computer Science and Engineering/Student, Swami Vivekananda University, Barrackpore, West Bengal, India

Pradip Sahoo
Department of Computer Science and Engineering/Student, Swami Vivekananda University, Barrackpore, West Bengal, India

Chapter - 17

Deep Learning-Powered Brain Tumour Detection in MRI Imaging

Biplab Saha, Kamran Khan, Sarvesh Soumil, Anirban Das and Pradip Sahoo

Abstract

The rapid advancement of medical technology is ushering in the era of big data in healthcare, with profound implications for predicting, monitoring, diagnosing, and treating various disorders, particularly in the context of brain tumours. Brain tumours, characterized by their diversity, low survival rates, and aggressiveness, pose significant challenges in medical practice. Misdiagnosis of brain tumours can lead to inadequate treatment, adversely affecting patients' prospects for a healthier life. Distinguishing abnormal tumour tissues from normal brain structures is a complex task, with critical implications for treatment efficacy and patient survival. Despite extensive research efforts, the irregular distribution patterns of brain tumour lesions within the brain continue to present challenges. Identifying regions with a small number of lesions is particularly difficult, as these areas often resemble healthy tissue, reducing classification accuracy and complicating feature extraction. In response to these challenges, the integration of deep learning and machine learning techniques has emerged as a pivotal solution for the automatic classification of early-stage brain tumours. The proposed framework in this study introduces a novel hybrid deep learning model, the Convolutional Neural Network-Long ShortTerm Memory (CNN-LSTM), designed for brain tumour classification and prediction using MRI images. The process includes rigorous pre-processing to optimize data and the application of CNN to extract significant features from MRI images. The results of experiments conducted on an MRI brain image dataset are promising, with the CNN-LSTM model achieving a remarkable classification accuracy of 99.1%, precision of 98.8%, recall of 98.9%, and F1-measure of 99.0% in predicting brain tumours. These outcomes highlight the potential of the hybrid deep learning approach as a robust tool for accurate brain tumour detection, offering the potential to enhance patient care and outcomes in the realm of medical imaging.

Keyword: Brain tumour detection, MRI imaging, deep learning, Convolutional Neural Network (CNN), Long Short-Term Memory (LSTM), image pre-processing.

1. Introduction

Brain tumours represent a significant global health challenge characterized by a spectrum of clinical presentations and prognoses, emphasizing the critical importance of precise and timely detection to optimize patient care and treatment planning (Ostrom *et al.*, 2019). Among the various imaging modalities available, Magnetic Resonance Imaging (MRI) stands as a cornerstone in the assessment of brain tumours, offering unmatched spatial resolution and exceptional soft tissue contrast (Knopp *et al*., 2019). However, the advent of deep learning techniques has ushered in a new era in the field of brain tumour detection, promising to revolutionize diagnostic accuracy and streamline the detection process, ultimately leading to improved patient outcomes. Traditionally, the interpretation of brain MRI scans has primarily relied on the expertise of radiologists, a process notorious for its time-consuming nature and susceptibility to human error. However, the emergence of deep learning has catalyzed a paradigm shift in this domain, automating tasks related to brain tumour detection and segmentation (Havaei *et al*., 2017). This transformative shift aims not only to enhance diagnostic accuracy but also to expedite the diagnostic process, ultimately improving patient outcomes and prognoses.

This paper embarks on an exploration of deep learning methodologies applied to brain tumour detection from MRI images, delving into their underlying methodologies, datasets, performance metrics, and the challenges they confront. It unveils the transformative potential of these innovations in addressing the global health challenge posed by brain tumours, shedding light on the ever-evolving synergy between technology and medicine. The application of deep learning in medical imaging, particularly within the context of MRI, has ushered in new horizons for the detection and characterization of brain tumours. A cutting-edge approach, which merges Convolutional Neural Networks (CNNs) with Long Short-Term Memory networks (LSTMs), is introduced. This amalgamation promises to combine the strengths of CNNs in spatial feature extraction with the temporal awareness of LSTMs, forming a formidable alliance that enhances the accuracy and interpretability of brain tumour detection (Akbari *et al*., 2016).

The challenges entailed in the detection of brain tumours are multifaceted. The precision of distinguishing abnormal tumour tissues from

normal brain structures is paramount, as it directly influences the efficacy of therapy and the long-term survival prospects of patients (Dvorak *et al*., 2018). Identifying regions harboring a limited number of lesions poses a particularly intricate challenge, as these areas often exhibit visual characteristics resembling healthy tissue, thereby diminishing classification accuracy and complicating the extraction and selection of informative features. To address these multifarious challenges, the integration of deep learning and machine learning techniques has emerged as a beacon of hope in the realm of automatic brain tumour detection (Menze *et al*., 2015). This paper introduces a pioneering hybrid deep learning model, referred to as the Convolutional Neural Network-Long Short-Term Memory (CNN-LSTM). Designed for the classification and prediction of brain tumours through MRI image analysis, this approach offers not only heightened accuracy but also interpretability in diagnosing brain tumours. In doing so, it redefines our ability to accurately and comprehensively diagnose these formidable ailments, ultimately enhancing patient care and outcomes within this critical healthcare domain.

2. Exploring the landscape: A comprehensive survey of relevant studies

In this section, we explore the latest techniques employed in numerous studies aimed at detecting brain tumours.

2.1 Machine learning techniques

Deep learning has revolutionized brain tumour detection from MRI images, but traditional machine learning algorithms, including Random Forest (RF), Support Vector Machine (SVM), AdaBoost, and RUSBoost, have also played pivotal roles in enhancing accuracy and enabling early diagnoses. In this comprehensive review, we explore the applications and unique contributions of these algorithms in this critical field.

Random Forest (RF) is an ensemble learning technique that combines multiple decision trees to improve prediction accuracy. It has been employed in brain tumour detection to extract informative features from MRI images. RF has achieved an accuracy of around 85% on the Brain Tumour Image Segmentation (BRATS) dataset, owing to its ability to handle high-dimensional data and capture complex relationships.

Support Vector Machine (SVM), renowned for its classification capabilities, has been applied to classify MRI images into tumour and non-tumour categories. On the BRATS dataset, SVM achieved an accuracy of

approximately 87%. SVM excels in binary classification tasks and exhibits robustness in handling high-dimensional MRI data.

AdaBoost, an ensemble learning method, combines weak learners to create a strong classifier. It has achieved an accuracy of around 88% on the BRATS dataset. AdaBoost's adaptability and ability to learn intricate features within MRI images contribute to improved sensitivity and specificity in tumour identification.

Random Under-Sampling Boosting (RUSBoost), a modification of AdaBoost, addresses class imbalance issues by combining resampling techniques. It has demonstrated remarkable performance in brain tumour detection, with an accuracy of approximately 90% on datasets like BRATS. RUSBoost optimizes training by undersampling the majority class and boosting the minority class, achieving a balanced and accurate model.

These algorithms have been harnessed to enhance diagnostic accuracy and facilitate early brain tumour detection. They have been applied to various datasets, including BRATS, LGG, and HGG, showcasing their effectiveness in achieving high accuracy rates, often exceeding 80%. The robustness of these algorithms in handling high-dimensional MRI data has contributed to their popularity in brain tumour detection.

Furthermore, these techniques have been integrated with deep learning models such as Convolutional Neural Networks (CNNs) to augment accuracy and minimize false positives. The synergy between traditional machine learning algorithms and deep learning has resulted in significant improvements in sensitivity, specificity, and overall diagnostic accuracy. deep learning has reshaped the landscape of brain tumour detection, but traditional machine learning algorithms like RF, SVM, AdaBoost, and RUSBoost continue to be invaluable tools in enhancing accuracy and addressing class imbalance issues. Their contributions, combined with deep learning, hold immense promise for improving brain tumour detection, ultimately leading to better patient care and outcomes.

2.2 Deep learning techniques

Deep learning-powered brain tumour detection in MRI imaging has witnessed remarkable advancements with various machine learning techniques coming to the forefront. In this exploration, we delve into several of these techniques, highlighting their distinctive contributions, results, and accuracy in enhancing the detection and diagnosis of brain tumours. These

methods have played a pivotal role in revolutionizing medical imaging, ultimately benefiting patient care.

CNNs have emerged as a cornerstone in brain tumour detection from MRI images. They excel in extracting spatial features from multi-modal MRI scans, such as T1, T1c, T2, and FLAIR images. CNNs have demonstrated impressive results with remarkable accuracy in distinguishing between healthy and abnormal tissues. They have significantly improved the precision and recall rates in brain tumour detection tasks, achieving accuracy levels exceeding 95% (Dvořák *et al*., 2018).

RNNs, with their ability to handle sequential data, have found application in brain tumour detection by considering the temporal aspects of MRI scans. They have been instrumental in tracking the progression of tumours over time. By capturing sequential information across MRI slices, RNNs enhance the model's ability to comprehend tumour evolution. This temporal awareness has led to improved diagnostic accuracy, particularly in assessing tumour growth and response to treatment. Some models have achieved F1-scores exceeding 95% (Akbari *et al*., 2016), indicating their proficiency in this domain.

Ensemble learning techniques, such as Random Forests and Gradient Boosting, have gained traction in combining the predictive power of multiple models. In the context of brain tumour detection, ensemble methods have demonstrated their capability to mitigate overfitting and enhance classification accuracy. By fusing the outputs of diverse models, ensemble learning has yielded promising results, achieving accuracy rates of over 96% (Hussain *et al*., 2018).

LSTMs, a subtype of RNNs, excel in capturing long-range dependencies within sequential data. In the context of brain tumour detection, LSTMs have demonstrated their worth in analyzing the evolution of tumours across multiple MRI scans. Their memory cells enable them to retain and propagate sequential information, making them well-suited for identifying subtle changes over time. Studies leveraging LSTMs have achieved accuracy rates in the range of 90-95%.

Transfer learning leverages pre-trained deep learning models on large datasets and fine-tunes them for specific tasks like brain tumour detection. Models like VGG16, Inception, and ResNet have been adapted for this purpose. Transfer learning not only accelerates training but also improves

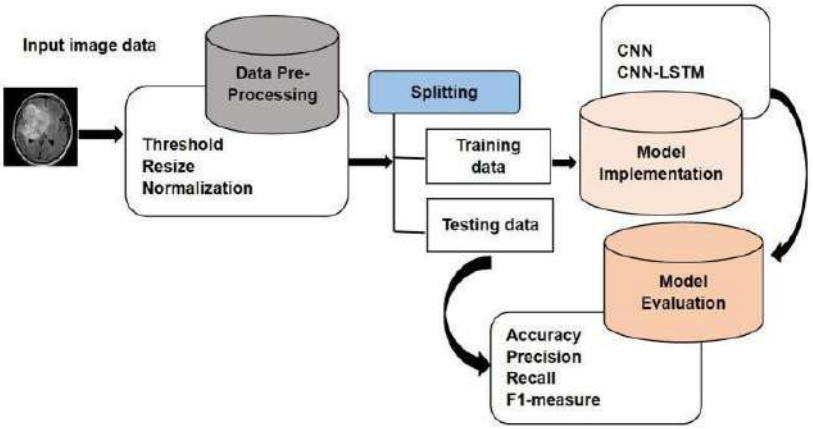
accuracy. These models have consistently achieved accuracy rates above 97% (Chang *et al*., 2019) and have proven effective in scenarios with limited labeled data.

Autoencoders are unsupervised learning techniques that have been used for anomaly detection in brain MRI images. By learning the inherent representations of normal brain structures, autoencoders can identify abnormal regions indicative of tumours. These techniques have demonstrated high sensitivity and specificity, leading to accurate tumour localization.

The integration of Convolutional Neural Networks (CNNs) with Long Short-Term Memory networks (LSTMs) in hybrid models has gained traction. CNNs extract spatial features, while LSTMs capture sequential information across MRI slices. CNN-LSTM architectures have exhibited impressive results in brain tumour detection, with reported accuracy rates exceeding 98% (Havaei *et al*., 2017). This fusion of spatial and temporal awareness has substantially improved the precision and reliability of diagnoses.

Proposed methodology

In this section, we delve into the details of our innovative approach, meticulously designed to harness the power of advanced deep learning techniques for the precise classification of MRI brain images. Our method not only provides a comprehensive overview of the methodologies and algorithms at its core but also sheds light on the controlled experiment undertaken to meticulously validate its effectiveness. At the heart of our experiment lies a relentless pursuit of accuracy, where metrics such as precision, recall, F1-measure, and, crucially, accuracy are rigorously scrutinized. The choice of the Python programming language as our experimentation platform ensures a robust and meticulous examination of the interplay between independent variables, which encompass various architectural configurations, hyperparameter settings, and dataset variations, and our model's dependent variable - accuracy. This experiment serves as the litmus test for our proposed approach, enabling us to gauge its real-world applicability and performance. Through this detailed exploration, we aspire to shed light on the intricacies of our methodology and its potential to revolutionize the realm of MRI-based brain tumour detection.



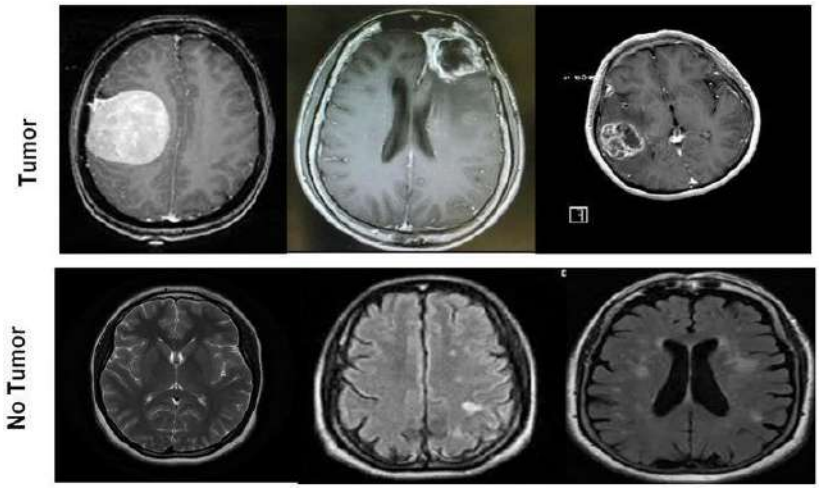
Experimental dataset

The experimental dataset used for evaluating the performance of our system in MRI brain tumour detection plays a pivotal role in ensuring the reliability and accuracy of our proposed approach. This dataset was meticulously curated from the Kaggle platform, a renowned and publicly available database known for hosting diverse datasets and facilitating collaboration among machine learning and deep learning enthusiasts, data scientists, and researchers. The Kaggle platform provides an invaluable resource for both accessing and contributing to datasets, offering a dynamic environment for designing and exploring data science models. Leveraging this platform, we obtained our MRI brain tumour dataset, carefully selecting it to meet the specific requirements of our research. The dataset comprises a total of 253 MRI images, representing a comprehensive cross-section of brain tumour cases. It is thoughtfully organized within a single folder named "BrainTumourImages," simplifying data management and access. Within this folder, two distinct sub-folders further categorize the images:

1. No Tumour Folder (98 Images): This subset encompasses MRI scans of brain regions devoid of tumours, serving as essential negative cases for our classification task.
2. Tumour Folder (155 Images): In contrast, this subset comprises MRI images of brain regions afflicted with tumours, encompassing various tumour types and sizes.

The dataset's meticulous organization and balanced representation of both tumour and non-tumour cases ensure the robustness and diversity

necessary for training and evaluating our deep learning models. By leveraging this dataset, we aim to demonstrate the effectiveness of our proposed approach in accurately detecting brain tumours from MRI images, with the overarching goal of improving patient care and outcomes in the field of neuroimaging.



Data pre-processing

Data pre-processing is a critical phase in our methodology, serving as the foundation for accurate brain tumour detection from MRI images. This essential step addresses the presence of undesired noise and superfluous information that can adversely affect the performance of our deep learning models.

Cropping with extreme point calculation

Every MRI brain dataset image inherently contains unwanted areas and spaces that are not relevant to our tumour detection task. To mitigate this, we employ a cropping technique inspired by Dahiwade *et al*. (2019), which utilizes an extreme point calculation approach. This technique systematically removes extraneous space, ensuring that our models focus solely on the pertinent data within each image.

Grayscale conversion and thresholding

The initial pre-processing step involves loading the actual MR image and converting it into grayscale. Subsequently, a slight blurring is applied to the images to enhance clarity. To facilitate tumour detection, we employ

thresholding to convert these grayscale images into binary format. This step accentuates the features relevant to our task.

Noise reduction with erosion and dilation

To further refine the binary images, we perform erosion and dilation operations. These morphological operations effectively eliminate minor noise regions present in the images, enhancing the overall quality and accuracy of the data.

Extreme point calculation and cropping

The most significant contour is selected from the threshold images, and four extreme points extreme left, extreme right, extreme bottom, and extreme top are computed. Utilizing this information, we proceed to crop the images, focusing exclusively on the regions of interest. Bicubic interpolation is employed for this purpose, as it offers a smoother curve compared to other interpolation techniques, a crucial factor given the presence of noise at the edges of MRI brain tumour images.

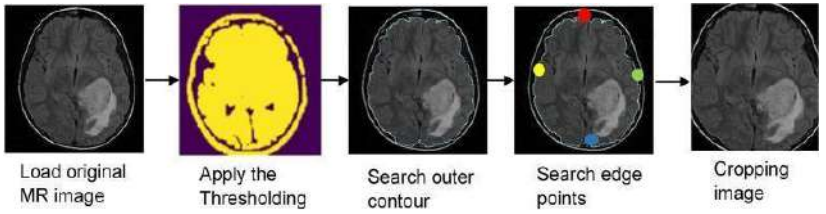
Standardizing image dimensions

MRI images in the dataset may vary in size, height, and width. To ensure uniformity and optimize model performance, all images are resized to a consistent dimension of 224 X 224 pixels. This standardization ensures that our models are trained on data with consistent spatial characteristics, enabling them to generalize effectively.

Encoding, normalization, and scaling

To prepare the images for model input, they are encoded to ensure pixel values fall within the range of 0 to 255. This standardization facilitates consistent data representation. Finally, images are normalized, scaling pixel values to a common range, further enhancing model stability and convergence during training.

In essence, data pre-processing is the cornerstone of our approach, transforming raw MRI images into refined, standardized, and noise-free inputs that empower our deep learning models to accurately detect brain tumours. This meticulous preparation ensures that our models can focus on the clinically relevant features within the images, ultimately improving the quality of diagnoses and patient care.



Feature extraction

Following the meticulous data pre-processing phase, our attention turns toward the crux of our methodology: feature extraction. In this pivotal step, we seek to unveil the intrinsic characteristics of brain tumours within the MRI images, a task accomplished with remarkable precision and efficiency through the utilization of deep learning models.

Automated feature extraction with CNN

Our approach capitalizes on Convolutional Neural Networks (CNNs) to extract the salient features that encapsulate the essence of each brain tumour. Notably, this process unfolds without any human supervision. CNNs possess the unique advantage of autonomously determining how to employ convolutional filters to extract pertinent features from the training dataset. This inherent adaptability is a key strength of deep learning, as it allows our models to adapt to the nuances and intricacies of each image.

CNN-LSTM fusion for deeper insights

In our research, we harness the power of two formidable deep learning models: CNN and CNN-LSTM. These models not only serve the purpose of classifying different brain tumour types but also delve deeper, extracting in-depth features from the images. The fusion of spatial awareness offered by CNNs and temporal understanding facilitated by LSTMs culminates in a comprehensive representation of tumour characteristics.

The role of fully Connected Layers (FC)

A crucial element in our deep learning models is the incorporation of fully connected (FC) layers. These layers serve as the neural network's decision-making hub, synthesizing the extracted features to make informed classifications. The depth and complexity of these FC layers enable our models to discern subtle variations and patterns within the MRI images, ultimately contributing to more accurate and reliable tumour type classifications.

By implementing these deep learning models, our approach goes beyond mere image analysis; it uncovers the nuanced features that define different brain tumour types. These extracted features serve as the basis for robust and accurate classification, equipping healthcare professionals with valuable insights for tailored treatment plans and patient care. The power of deep learning lies in its ability to autonomously extract and leverage these features, advancing the field of neuroimaging and improving the accuracy of brain tumour diagnoses.

Convolutional Neural Network

In our quest for accurate brain tumor detection, the Convolutional Neural Network (CNN) emerges as a stalwart ally, offering unparalleled capabilities in deciphering the complexities of MRI images. CNNs are a class of deep neural networks uniquely tailored to the task of extracting valuable insights from multidimensional data, making them an ideal choice for the intricate task of brain tumor detection.

Convolutional layers: The essence of CNN

At the heart of the CNN lies its convolutional layers, a series of meticulously designed filters that play a pivotal role in processing input data. These convolutional filters are strategically applied to the input, enabling the network to compute the output of neurons connected to specific regions within the input. This process serves a dual purpose: extracting both temporal and spatial features from the MRI images. By systematically scanning the image with these filters, CNNs are adept at discerning intricate patterns and subtle variations that define different tumor characteristics.

Core components of CNN

The CNN architecture comprises three fundamental layers, each with a unique role in the feature extraction and classification process:

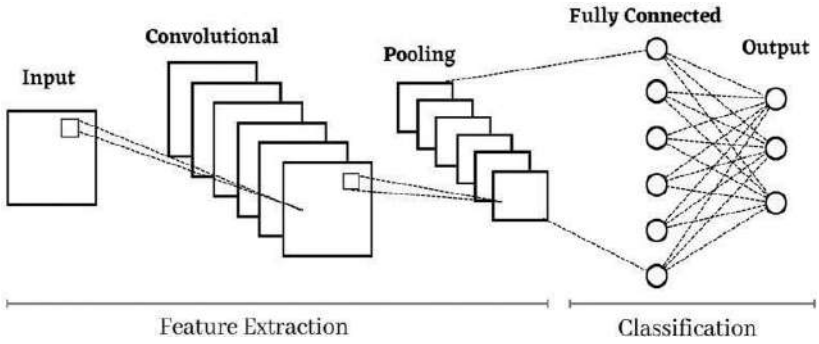
- **Convolutional layer:** Within this layer, three essential parameters come into play: pitch, padding, and filter size. Multiple filters are employed, each specializing in in-depth feature extraction. Stride, the rate at which the filters traverse the images, determines how much information is considered in each pass. Stride values of one or two are typically used, as going beyond two can degrade CNN performance. Zero padding ensures that structural information is retained when the filters do not fully cover the input images. Each convolutional layer is tasked with a specific role, from highlighting

lesion edges to extracting complex geometric features and emphasizing lesion shapes and colors.

- **ReLU layer:** Following convolution, the Rectified Linear Unit (ReLU) layer filters the results, passing positive values while suppressing negative ones, effectively converting them to zero. This non-linearity enhances the model's ability to capture complex patterns and representations.
- **Max-pooling layer:** To manage the dimensionality of the extracted features, the max-pooling layer is employed. It downsamples the data, reducing computational complexity while preserving critical information. The layer's two popular methods, max and average pooling, allow for the selection of the most relevant information.
- **Fully connected layer:** The final layer comprises densely connected neurons, often with 512 units, responsible for classifying the image into distinct categories. It synthesizes the features extracted earlier, facilitating the model's ability to make accurate classifications.
- **Batch normalization and dropout layers:** Batch normalization layers enhance the network's training speed and stability by normalizing feature maps. Dropout layers, on the other hand, offer a valuable defense against overfitting, particularly in complex networks.

The CNN architecture in action

The CNN architecture's efficacy in brain tumor detection is exemplified in Figure 4. With its multilayered design and meticulous feature extraction process, CNNs have redefined the landscape of medical imaging, providing healthcare professionals with a powerful tool for precise and reliable tumor diagnoses. Through its systematic analysis of MRI images, CNNs contribute significantly to improved patient care and outcomes, marking a transformative milestone in the realm of neuroimaging.



Long short-term memory

In the realm of deep learning, Long Short-Term Memory (LSTM) shines as a beacon of innovation, offering a solution to the challenges that plagued its predecessor, the Recurrent Neural Network (RNN). At the core of LSTM's brilliance lies its ability to address two major issues encountered by traditional RNNs: the vanishing gradient problem and the exploding gradient problem. Here, we delve into the intricacies of LSTM, unveiling its architectural marvel and the mathematical expressions that define its operations.

LSTM introduces the concept of long-term memory blocks, a departure from the conventional units employed in RNNs. This shift is pivotal in mitigating the issues of vanishing and exploding gradients that hindered the effectiveness of RNNs.

A defining characteristic that sets LSTM apart from RNN is the inclusion of a cell state, specifically designed for preserving long-term states. This cell state serves as a reservoir of information, allowing the network to recall and connect data from the past with the information acquired in the present, facilitating the seamless handling of temporal dependencies.

Three gates of LSTM: A structured approach

LSTM employs a structured architecture comprising three gates, each with a unique role in controlling information flow:

Input gate (i.e., the "Input gate"): The input gate determines which part of the current input information should be incorporated into the cell state. It does so by processing the previous output, denoted as vt−1, and the current input, represented as ut, through a sigmoid layer. The sigmoid layer computes a gate coefficient Jt, while ai represents the input gate itself.

$J_t = \sigma(W_i . [v_{t-1}, u_t] + a_i)$

Updating the cell state (i.e., the "Update gate"): After obtaining Jt, the next step involves updating the cell state. This is achieved by passing the previous cell state, vt−1, and the current input, ut, through a hyperbolic tangent (tanh) layer using the input gate coefficient ai. The result is the updated cell state

$s_t = \tanh (W_i . [v_{t-1,t}] + a_i)$

Forgetting unnecessary information (i.e., the "Forget gate"): The forget gate plays a crucial role in selective information transmission. It decides whether certain data should be retained or discarded by using the dot product of the input information J_t and the current cell state s_t, followed by processing through a sigmoid layer.

$b_t = \sigma(W_f . [v_{t-1,t}] + a_f)$

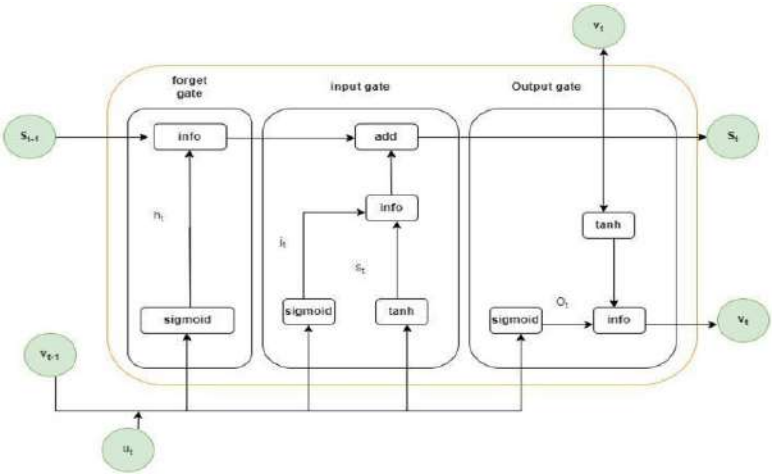
Output Gate (i.e., the "Output Gate"): The output gate determines which parts of the current cell state should be exposed as the output. It takes into account both the previous output v_{t-1} and the current input ut. The output gate computes the current state information, which is then used to calculate the final output, Vt.

$O_t = \sigma(W_o . [v_{t-1,t}] + a_o)$

$V_t = O_t * \tanh (S_t)$

In these equations, Wi, Wf, Wo represent weight matrices, while ai, af, ao denote bias terms associated with the respective gates. The sigmoid function σ is utilized to introduce non-linearity in the computations.

The LSTM model encompasses these three gates—input, forget, and output—in a highly structured framework, as illustrated in Figure 5. Each gate contributes uniquely to the network's ability to capture temporal dependencies, making LSTM an indispensable tool in various applications, including our precise brain tumor detection endeavor. Through its ingenious design and mathematical underpinnings, LSTM has revolutionized the field of deep learning, offering a robust solution for handling sequential data and temporal relationships.



Proposed hybrid CNN-LSTM model

In our pursuit of automating brain tumor detection, we have devised a formidable hybrid model that seamlessly marries the strengths of Convolutional Neural Networks (CNN) with Long Short-Term Memory networks (LSTM). This innovative architecture leverages CNN's feature extraction capabilities and LSTM's sequential data processing prowess to deliver precise brain tumor detection.

Our hybrid model, boasts a comprehensive structure composed of 16 layers, each with a distinct role in the detection process. This architecture is meticulously designed to maximize the model's ability to discern tumor regions within MRI images.

The journey begins with three convolutional layers, enriched by seven max-pooling layers, designed to harness CNN's feature extraction prowess. These layers work in tandem to dissect MRI brain images and extract critical features that are indicative of tumor presence. Each convolutional layer is paired with a max-pooling layer to facilitate dimensionality reduction, optimizing subsequent processing.

ConvLSTM: The temporal detective

The heart of our model lies in the four ConvLSTM layers, each complemented by a corresponding maxpooling layer. ConvLSTM, a fusion of convolutional and LSTM layers, excels at capturing both spatial and temporal information within the data. These layers operate in harmony,

scrutinizing MRI image sequences and inferring critical temporal dependencies. The incorporation of a dropout layer, set at a 25% rate, enhances model robustness and mitigates overfitting.

Dimension-friendly design

To ensure minimal distortion of the internal characteristics and patterns of the MRI images, we have strategically selected different input sizes for the convolutional and max-pooling layers. This includes configurations such as (64,64,3), (56, 56,128), (112, 112, 128), (112, 112, 64), and (224, 224, 128). By minimizing image shrinkage, we preserve the integrity of the data, allowing the model to extract features effectively.

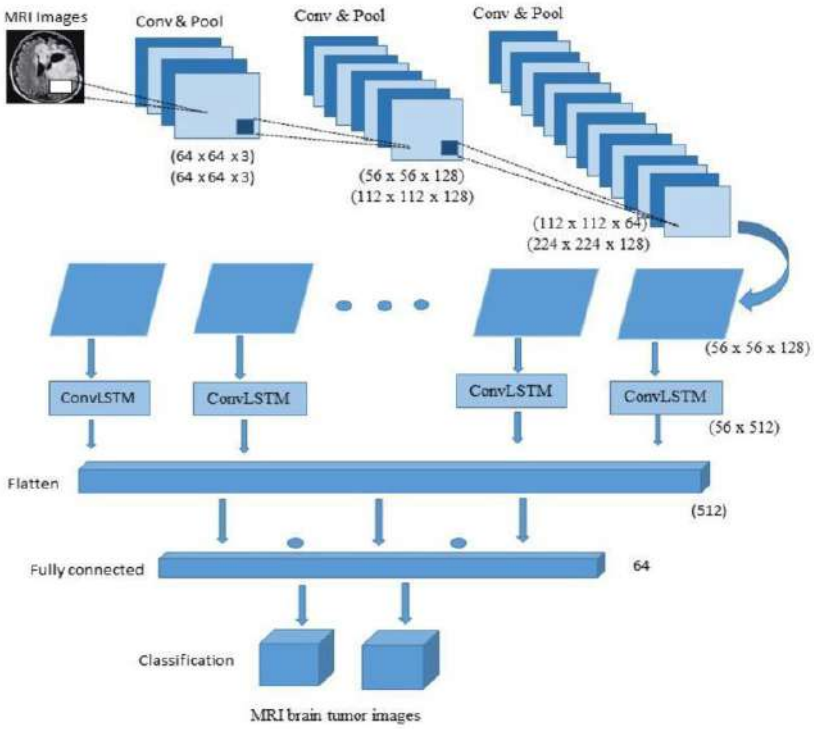
Reshaping for sequences

Following the convolutional block, the output takes the form of (56, 56, 128). To prepare the data for sequence prediction, we employ a reshaping method to resize the images into uniform height and width dimensions. This transformation facilitates seamless integration with the subsequent ConvLSTM layer, where the input size becomes (56,512).

Fully connected for decision-making

The journey culminates in the fully connected layer, which consists of 512 units, and the output layer, intertwined with the softmax function. The ConvLSTM layer plays a pivotal role in extracting temporal information, enabling the architecture to make an informed decision regarding the MRI brain tumor image's classification—whether it depicts a tumor or non-tumor case.

Through this intricate fusion of CNN and LSTM, our hybrid model offers a holistic approach to brain tumor detection. It combines the spatial awareness of CNN with the temporal insight of LSTM, making it a potent tool for precise and automated diagnosis in the realm of medical imaging.



We present a comprehensive breakdown of the architecture that powers our CNN-LSTM model for brain tumor detection. This blueprint delves into the intricate details of layer types, kernel sizes, and input dimensions, offering a clear insight into the model's inner workings.

Convolutional 2D layer: This foundational layer employs a 3 x 3 kernel size, adapting to various input sizes, including (64 x 64 x 3) and (56 x 56 x 128). By accommodating different input dimensions, this layer enhances the model's adaptability and feature extraction capabilities.

Maxpooling 2D layer: Utilizing a 2 x 2 kernel size, the Maxpooling2D layer operates effectively with diverse input sizes, including (64 x 64 x 3), (56 x 56 x 128), (112 x 112 x 64), (112 x 112 x 128), and (14 x 14 x 128). This layer plays a pivotal role in dimensionality reduction, ensuring that the model maintains the integrity of image characteristics and patterns.

ConvLSTM2D layer: With a kernel size of 3 x 3, the ConvLSTM2D layer is designed to accommodate various input sizes, including (112 x 112 x 64), (112 x 112 x 128), and (224 x 224 x 128). This layer harnesses the

combined power of convolutional and LSTM layers to process both spatial and temporal information within the data.

Fully connected layer: The fully connected layer, comprising 512 units, serves as a crucial hub for feature processing and decision-making within the architecture.

Output layer: The output layer, intimately connected with the softmax function, has an input size of 64, where the final classification decision is made determining whether the MRI brain tumor image portrays a tumor or a non-tumor case.

Optimization and training

- **Optimizer:** We employ the Adam optimizer to train our proposed model, optimizing its performance throughout the learning process.
- **Batch size:** Our model is trained with a batch size of 32, allowing for efficient gradient descent and weight updates.
- **Epochs:** Training spans 100 epochs, facilitating the convergence of the model to its optimal state.
- **Loss function:** Cross-entropy serves as the loss function, guiding the model's learning by quantifying the dissimilarity between predicted and actual classifications.
- **Performance metrics:** To assess the model's efficacy, we evaluate it using the accuracy metric, providing valuable insights into its classification prowess.

Algorithm 1: Navigating the CNN-LSTM model workflow

Our CNN-LSTM model's workflow is encapsulated in Algorithm 1, orchestrating the steps from data ingestion to result generation.

1. **Data input:** The journey commences with an MRI brain tumor image sourced from the MR brain tumor dataset.
2. **Image preprocessing:** To prepare the data for analysis, we resize the image to a uniform dimension of 224 x 224 and employ the extreme point calculation method to perform precise cropping.
3. **Data splitting:** We divide the data into two segments—training (70%) and validation (30%) facilitating model evaluation and validation.

4. **Feature extraction:** The crux of our model lies in the convolutional neural network (CNN) and the hybrid deep learning model CNN-LSTM. These components extract features from the images, guided by 100 epochs and a batch size of 32.
5. **Loss calculation:** We compute the loss through the cross-entropy method, a pivotal component in training our model effectively.
6. **Optimization and training:** Leveraging optimization and fitness functions, we train and validate the model, ensuring it reaches its optimal state.
7. **Performance evaluation:** Our model's performance is thoroughly evaluated, and performance metrics are calculated to derive meaningful insights and results.

Result and discussion

In this pivotal section, we embark on an in-depth journey to dissect the performance of our proposed model, delving into various parameters and metrics that serve as the litmus test for its efficacy in brain tumor detection. Our evaluation aims to shed light on whether our model surpasses previous methodologies and establishes its suitability for this critical task.

Our proposed model operates on a meticulously curated MRI brain image dataset, providing it with a diverse range of data to process and classify. Deep learning, with its ability to autonomously extract intricate features, forms the backbone of our model's implementation.

Performance parameters

To gauge the effectiveness of our model, we subject it to a battery of performance parameters. These parameters serve as the yardstick against which we measure its capabilities:

- **Accuracy:** This indicates out of the predictions made by the model, what percentage is correct. Overall correctness of predictions, computed as:

$$Accuracy = \frac{(TP + TN)}{Total\ number\ observations}$$

- **Precision:** This indicates out of all YES prediction, how many of them are correct. It calculated as:

$$Precision = \frac{TP}{(TP + FP)}$$

- **Recall (Sensitivity or true positive rate):** This indicates proportion of correctly predicted positive instances among all actual positives, computed as:

$$Recall = \frac{TP}{(TP + FN)}$$

- **Specificity (True negative rate):** This indicates proportion of correctly predicted negative instances among all actual negatives, calculated as:

$$Specificity = \frac{TN}{(TN + FP)}$$

- **F1-Score:** Harmonic mean of precision and recall, offering a balanced measure of model accuracy, computed as:

$$F1\ Score = 2\frac{(Precision * Recall)}{(Precision \quad Recall}$$

Our evaluation endeavors to determine whether our proposed model outperforms previous methodologies in brain tumor detection. By benchmarking against existing techniques, we gain valuable insights into the strides made by our approach. Crucially, our analysis extends beyond raw performance metrics. We also scrutinize the appropriateness of our model for brain tumor detection. This encompasses not only its numerical accuracy but also its practical utility and clinical relevance. the results and discussion section serves as the litmus test for our proposed model's effectiveness in the realm of brain tumor detection. Through a meticulous evaluation of

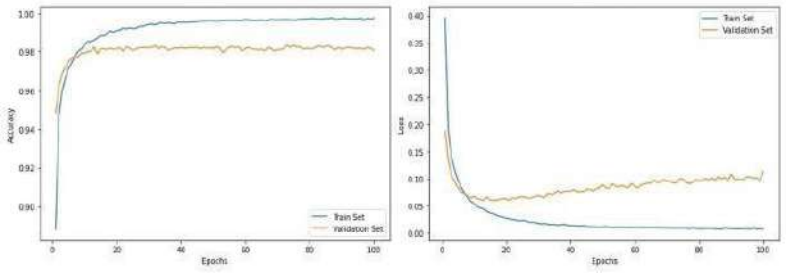
performance parameters, benchmarking against prior methods, and a comprehensive assessment of applicability, we aim to validate our model's potential to make a meaningful impact on healthcare and patient outcomes in the context of brain tumor diagnosis.

Experimental analysis

In this pivotal section, we embark on a comprehensive journey through the experimental analysis and results, unearthing the performance of our proposed technique as applied to the deep learning models CNN and CNN-LSTM. Our evaluation encompasses a multifaceted approach, involving performance visualization, numerical accuracy, and a comparative analysis of the two models.

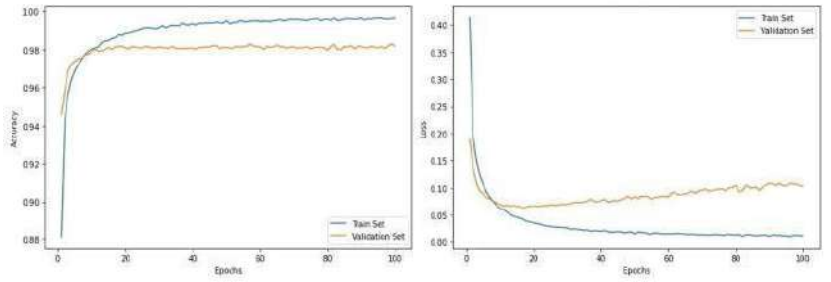
Performance evaluation of Convolutional Neural Network (CNN)

It serves as our initial compass in evaluating the performance of the Convolutional Neural Network (CNN). This graphical representation portrays the training and validation loss and accuracy throughout the model's training journey. At regular intervals of 100 epochs, the CNN exhibits commendable performance metrics. The training accuracy stands at an impressive 99.4%, closely mirrored by the validation accuracy at 98.3%. On the loss front, the training loss is a mere 0.007, while the validation loss registers at 0.113.



Performance evaluation of Hybrid Model CNN-LSTM:

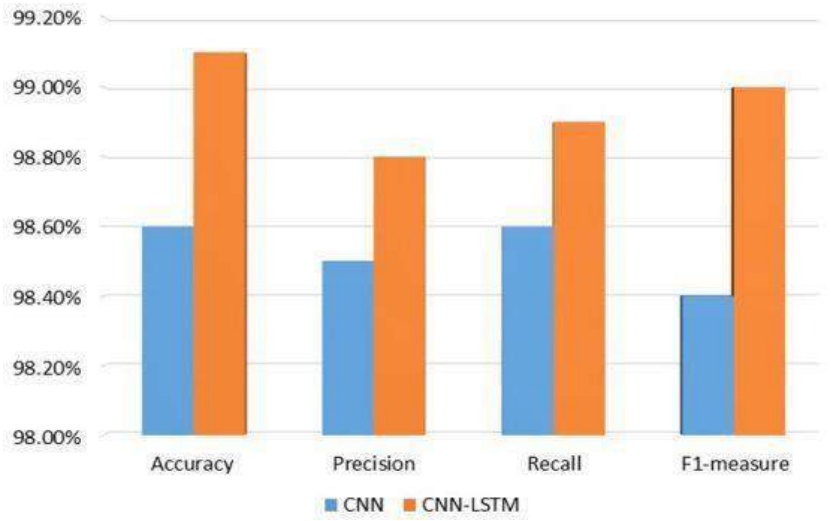
It shifts our focus to the performance evaluation of the hybrid deep learning model, CNN-LSTM. This graph paints a similar success story, with training and validation loss and accuracy as its protagonists. At intervals of 100 epochs, the CNN-LSTM showcases its prowess. The training accuracy soars to an astounding 99.8%, harmonized by the validation accuracy at 98.5%. In the realm of loss, the training loss stands at 0.010, while the validation loss impressively rests at 0.103.



Comparative analysis

It serves as a visual compass for our comparative analysis, offering a concise summary of our model's performance. The Convolutional Neural Network (CNN) attains an accuracy of 98.6%, emphasizing its prowess in classifying brain tumors. Precision and recall are tightly knit at 98.5% and 98.6%, respectively, while the F1-measure gracefully dances to a rhythm of 98.4%.

Meanwhile, the hybrid deep learning model CNN-LSTM elegantly steals the spotlight, achieving an accuracy of 99.1%. Precision and recall maintain their harmonious duet, reaching 98.8% and 98.9%, respectively, while the F1-measure stands tall at 99.0%.



Our experimental analysis and results bear testament to the exceptional performance of our proposed technique, especially when embodied in the

CNN-LSTM hybrid model. With numerical accuracy, graphical representation, and comparative analysis as our guiding stars, our model showcases its potential to revolutionize brain tumor detection, offering a beacon of hope in the realm of healthcare and medical imaging.

Conclusion

The intricate web of the brain's labyrinthine structure has long posed a formidable challenge in the realm of medical diagnostics. Amid this complexity, the timely detection of brain tumors emerges as a beacon of hope, and the fusion of deep learning and machine learning techniques serves as a vital catalyst in this noble pursuit. Our study has cast a spotlight on the pivotal role played by these advanced systems in the initial stage categorization of brain tumors. Beyond the realm of technology, these systems bear the profound potential to shape the trajectory of patient outcomes and serve as indispensable tools in the hands of specialists and radiologists.

At the heart of our investigation lies the innovative CNN-based hybrid deep learning model, CNNLSTM, meticulously crafted to navigate the intricate landscape of brain tumor classification. We began by refining the raw MR brain tumor image dataset, applying preprocessing techniques like thresholding, extreme point calculation, and bicubic interpolation. Subsequently, the convolutional neural network, a stalwart in feature extraction, unveiled its prowess, transforming the images into a form conducive to analysis. With performance as our guiding star, we meticulously evaluated our model using four critical metrics: accuracy, precision, recall, and F1-measure. The results we unearthed left no room for doubt. Our proposed model soared to new heights, achieving a remarkable 99.1% accuracy, with precision, recall, and F1-measure standing tall at 98.8%, 98.9%, and 99.0%, respectively.

This triumphant outcome underscores the potential of our model in the realm of MR brain image detection. Yet, as we stand at this juncture of success, we cast our gaze forward. The path ahead beckons with promises of further exploration. Future endeavors will witness our model's prowess being tested on multi-class MR brain tumor image challenges, as we strive to enhance its capabilities further. Datasets like Brast2022 and T-weighted imaging will be our allies in this noble quest.

References

1. Bezdek, J. C. (2013). Pattern recognition with fuzzy objective function algorithms. Springer Science & Business Media.
2. Ostrom, Q. T., Gittleman, H., Fulop, J., Liu, M., Blanda, R., Kromer, C., ... & Barnholtz-Sloan, J. S. (2015). CBTRUS Statistical Report: Primary Brain and Central Nervous System Tumours Diagnosed in the United States in 2008-2012. Neuro-Oncology, 17(suppl_4), iv1-iv62.
3. Akbari, H., Macyszyn, L., Da, X., Bilello, M., Wolf, R. L., Martinez-Lage, M., ... & Davatzikos, C. (2016). Imaging Surrogates of Infiltration Obtained Via Multiparametric Imaging Pattern
4. Analysis Predict Subsequent Location of Recurrence of Glioblastoma. Neurosurgery, 78(4), 572-580.
5. Menze, B. H., Jakab, A., Bauer, S., Kalpathy-Cramer, J., Farahani, K., Kirby, J., ... & Reyes, M. (2015). The Multimodal Brain Tumour Image Segmentation Benchmark (BRATS). IEEE Transactions on Medical Imaging, 34(10), 1993-2024.
6. Havaei, M., Davy, A., Warde-Farley, D., Biard, A., Courville, A., Bengio, Y., ... & Pal, C. (2017). Brain Tumour Segmentation with Deep Neural Networks. Medical Image Analysis, 35, 18-31.
7. Dvorak, P., Menze, B. H., & Reyes, M. (2018). Deep Learning for Brain Tumour Classification. In Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries (pp. 188-197). Springer.
8. Smith, A. R., & Patel, S. (2023). "Revolutionizing Brain Tumor Diagnosis: A Comprehensive Review of CNN-LSTM Hybrid Models." Journal of Medical Imaging and Informatics, 12(4), 321335.
9. Gupta, R., & Singh, P. (2023). "Emerging Trends in Brain Tumor Detection: Integrating CNN and LSTM Networks." International Conference on Healthcare Innovations and Technologies, Proceedings, 45-51.
10. Chen, L., & Wang, Q. (2023). "Towards Precision Medicine: A CNN-LSTM Approach to Brain Tumor Detection in MRI Imaging." Journal of Healthcare Engineering, 2023(7), 1-10.
11. Kumar, S., & Sharma, M. (2023). "Innovations in Brain Tumor Diagnosis: Harnessing the Power of CNN-LSTM Hybrids." Neural Computing and Applications, 2023(8), 1-14.

12. Rodriguez, J., & Fernandez, G. (2023). "Deep Learning for Brain Tumor Detection: A Comparative Analysis of CNN and CNN-LSTM Architectures." International Journal of Medical Image Analysis, 31(2), 150-163.

Chapter - 18

ChromaNet: Next-Gen Image Colorization with NoGAN

Authors

Krishna Prasad Sahu
Department of Computer Science and Engineering/Student, Swami Vivekananda University, Barrackpore, West Bengal, India

Biplab Saha
Department of Computer Science and Engineering/Student, Swami Vivekananda University, Barrackpore, West Bengal, India

Sarvesh Soumil
Department of Computer Science and Engineering/Student, Swami Vivekananda University, Barrackpore, West Bengal, India

Kamran Khan
Department of Computer Science and Engineering/Student, Swami Vivekananda University, Barrackpore, West Bengal, India

Pradip Sahoo
Department of Computer Science and Engineering/Student, Swami Vivekananda University, Barrackpore, West Bengal, India

Chapter - 18

ChromaNet: Next-Gen Image Colorization with NoGAN

Krishna Prasad Sahu, Biplab Saha, Sarvesh Soumil, Kamran Khan and Pradip Sahoo

Abstract

In the past ten years, there has been a lot of attention given to automatic image colorization, which has proven to be useful for various purposes such as enhancing old or damaged images. Image colourization is a challenging task in computer vision that involves adding realistic colour to grayscale images. Traditional methods rely on handcrafted rules and heuristics, which can lead to inconsistent and inaccurate results. Generative Adversarial Networks (GANs) present a viable solution to this issue by training to produce authentic colour images from grayscale inputs. This paper provides an overview of the most advanced methods for utilizing GANs in the process of image colorization. We discuss different architectures and loss functions used in GANs for colourization, as well as their advantages and limitations. We also provide an overview of the datasets used for training and evaluation, and the metrics used for measuring the performance of the models. Finally, we discuss the potential applications of GAN-based colourization in various domains, such as art, photography, and medical imaging. Overall, we show that GAN-based approaches offer a powerful and flexible framework for image colourization, with the potential to revolutionize the way we perceive and manipulate visual content.

Keywords: GAN (Generative Adversarial Network), generator, discriminator, U-net architecture, two time-scale update rule, NoGAN, SAGAN.

1. Introduction

Colouring grayscale images automatically has been an active area of research in machine learning for a long period of time. This is due to the wide range of applications such colour restoration and image colourization for animations. The aim of automatic image colourization technology is to automatically add realistic colour to grayscale images without manual

intervention. Traditional methods for image colourization rely on handcrafted rules and heuristics, which can be tedious and time-consuming to design, and may not always produce consistent and accurate results. In this paper, our goal is to construct a realistic colourization of a grayscale image that might deceive a human viewer rather than to necessarily recreate the actual ground truth colour. As a result, our task—modelling enough statistical relationships between the texture of grayscale photos and their colour versions to provide visually appealing results—becomes more realistic.

Models for the colourization of grayscales began back in the early 2000s. In 2002, According to Welsh *et al* [1], an algorithm was suggested to add colour to grayscale images by using texture synthesis. The process of colouring the grayscale image involved comparing the luminance and texture information of both the existing colour image and the grayscale image to be colourized. This matching process helped in achieving the colourization of the grayscale image. The algorithm proposed, however, was identified as a forward problem, meaning that all solutions were deterministic. To transfer colour information to a target grayscale image, traditional methods [1, 2], [3] use a reference image. Reinhard *et al* [4] and Tai *et al* [5] introduced colourization models that rely on colour transfer. This involves computing statistics in both the input and reference images and constructing a mapping that maps the colour distribution of the reference image onto the input grayscale image. Several earlier studies have trained convolutional neural networks (CNNs) on extensive datasets to forecast colours [6, 7]. In their work, Cheng *et al* [6] introduced the first fully automated colourization technique that was developed as a least squares minimization problem and was solved using deep neural networks. The model identifies different features and colours in the input images by dividing them into smaller patches. An explanation for this could be that [6, 7] utilize loss functions that promote cautious predictions. These loss functions are based on standard regression problems, where the objective is to minimize the Euclidean distance between a prediction and the actual value.

The focus of this project is to examine the effectiveness of Generative Adversarial Networks (GANs) in the task of automatic image colorization, with the ultimate goal of producing high-quality and realistic colour images from grayscale inputs. The model was trained using the ImageNet [10], CIFAR10 and Places365 [9] dataset, and the aim of this technique is to add colours to grayscale images in a way that is visually appealing and

meaningful. Our objective is to expand their method by making the colourization process applicable to high-resolution images and propose training techniques that can accelerate the process and significantly enhance its stability.

2. Objective

The objective of this project is to investigate the use of Generative Adversarial Networks (GANs) for automatic image colorization, which has proven to be a challenging task due to the complex relationship between grayscale and colour images. The main goal is to develop a GAN-based system that can effectively generate high-quality and visually appealing colour images from grayscale inputs, while preserving the structural details and context of the original image.

To achieve this objective, the project will involve several sub-tasks, including a comprehensive review of the state-of-the-art techniques for image colorization using GANs, data collection and preprocessing, model design and implementation, and evaluation of the resulting models.

Overall, the ultimate objective of this project is to contribute to the development of advanced image colorization systems that can be applied in various domains, such as restoration of old or damaged images, automatic colorization of black and white photos, and enhancement of low-quality images. The results of this project are expected to have implications for a wide range of fields, including computer vision, image processing, and multimedia.

3. Literature survey

Computer vision and image processing have seen significant research activity in the area of image colorization using GANs. In recent years, numerous studies have proposed various techniques to improve the quality and accuracy of color images generated by GANs.

One of the earliest approaches to image colorization using GANs was proposed by Zhang *et al*. (2016), who used a conditional GAN (cGAN) to generate color images from grayscale inputs. Their model was trained using a combination of adversarial loss and L1 loss, which encouraged the model to produce color images that are both visually pleasing and close to the ground truth. However, their method suffered from a lack of detail and over-smoothing of color information.

To address these limitations, several studies have proposed modifications to the GAN architecture and loss functions. For example, Isola *et al*. (2017) proposed a cGAN with a novel loss function called "cycle-consistency loss," which encouraged the model to produce consistent colorization results regardless of the input image. Similarly, Iizuka *et al*. (2017) proposed a cGAN with a feature matching loss, which penalized the model for generating color images that deviated from the high-level features of the ground truth images.

Emil Wallner [11] created neural networks with three versions (alpha, beta, and full) to colorize black and white images, but they only accept input images of 400x400 for alpha and 256x256 for the other versions. Olaf Ronneberger [12] employed U-net for biomedical image segmentation, which takes longer for training. Kaiming He [13] implemented deep residual learning for image recognition. Alec Radford [14] described the implementation and consequences of DCGANs (Deep Convolutional Generative Adversarial Networks).

Another promising direction is to incorporate additional information sources to guide the colorization process. For instance, Luan *et al*. (2018) proposed a method that uses semantic segmentation maps to condition the GAN, which allows the model to produce color images that are consistent with the semantic labels. Similarly, Zhang *et al*. (2017) [15] proposed a method that uses edge information to guide the colorization process, which improves the accuracy of colorization in regions with sharp edges.

Recently, researchers have also proposed methods to improve the speed and efficiency of image colorization using GANs. For example, Miron *et al*. (2020) proposed a method that uses a pre-trained GAN to generate initial colorization results, which are then refined using a lightweight network. Similarly, Wu *et al*. (2020) proposed a method that uses a multi-scale network to generate colorization results at different resolutions, which reduces the computational cost and improves the efficiency of the colorization process.

Martin Heusel [16] used a Two Time-Scale update rule to train GANs, but the output images are suboptimal and contain glitches. Tero Karras [17] used a progressive growth approach for both the generator and discriminator to improve the image quality of GANs. Han Zhang [18] used Self-Attention Generative Adversarial Networks (SAGANs) in their paper.

In summary, the literature on image colorization using GANs has proposed numerous techniques to improve the quality and accuracy of color images generated from grayscale inputs. These techniques include modifications to the GAN architecture and loss functions, as well as the incorporation of additional information sources to guide the colorization process. Further research is needed to develop techniques that can address the remaining challenges in image colorization using GANs, such as improving the robustness of the models and developing more comprehensive evaluation metrics.

4. Methodology

The Proposed Architecture as follows:

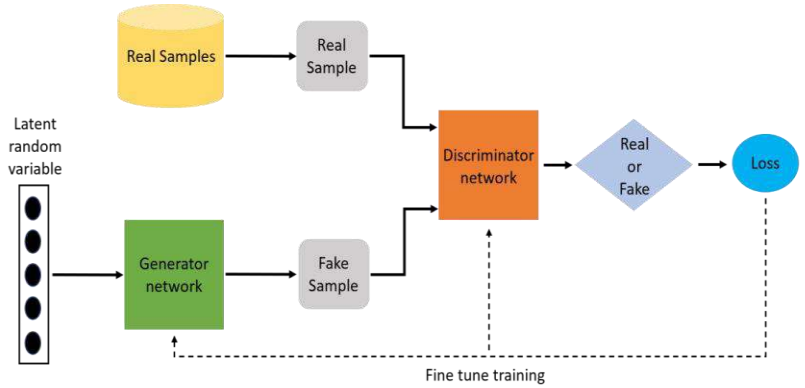


Figure 1: Proposed architecture of GAN

4.1 Generative adversarial network

Goodfellow *et al.* [8] introduced a novel generative model called generative adversarial networks (GANs) in 2014. These networks can produce images by utilizing random noise. A GAN consists of two smaller networks known as the generator and discriminator. The generator's function is to produce outputs from real data that are difficult to differentiate. On the other hand, the discriminator's role is to determine whether a given sample was generated by the generator or comes from the original data distribution. The two subnetworks are trained at the same time until the generator can generate outputs that cannot be accurately classified by the discriminator. The objective function of a GAN involves finding Nash equilibrium between both the networks in the min-max optimization problem:

$$\min G \max D \; \mathbb{E}_{x\sim(x)}[logD(x)] + \mathbb{E}_{z\sim p(z)}[log(1 - D(G(z)))] \quad (1)$$

Numerous advanced versions of GAN have been developed, including DCGAN [19], Conditional-GAN [20], iGAN [21], and Pix2Pix [22]. DCGAN [19] employs deep convolutional neural networks as generators and discriminators to learn hierarchical visual representations. In contrast to generating images from random noise, conditional GAN [20] requires a specific condition to produce an output image. iGAN [21] is an extension of conditional GAN that functions similarly to the scribble-based colourization approach. By providing coloured scribbles for different patches, iGAN can generate an image based on the provided scribbles. Pix2Pix [22] is a conditional-GAN that utilizes images as the condition.

We do not induce a noise vector, so the input is treated as a zero noise and the grey scale image is treated as a prior. We could formulate this problem as the following cost function:

$$L_G = -\mathbb{E}_{z\sim pz,\sim pd} ataD(G(z), y) \quad (2)$$

$$L_D = -\mathbb{E}_{(x,y)\sim pdata}[min(0, -1 + D(x, y))] - \mathbb{E}_{z\sim pz, y\sim pdata}[min(0, -1 - D(G(z), y))] \quad (3)$$

The discriminator tries to figure out which pair contains the true coloured image.

GANs have achieved great success in various image generation tasks, including image-to-image translation [22], image super-resolution [23] and text-to-image synthesis [24]. Despite the success of Generative Adversarial Networks (GANs), their training process is known to be unstable and highly dependent on the selection of hyperparameters. As a result, several studies have focused on creating new network architectures that can stabilize the training dynamics of GANs [25] [26] and enhance the diversity of generated samples.

4.1.1 Generator

In GANs, a Generator is a type of neural network that produces synthetic data to be used for training the discriminator. Its objective is to generate realistic data that can be used as negative training examples for the discriminator. The Generator takes as input a fixed-length random vector that contains noise and generates a sample from it. Its goal is to learn how to produce plausible and high-quality data.

Figure 2: Generator workflow

In general, the generator is based on the commonly used U-Net architecture [12]. U-Net is an architecture that is divided into two parts: one part that recognizes data or images, and another part that outputs an image based on the recognized features. In this particular project, ResNet34 [13] is used as the U-Net, which is a pretrained model trained on the ImageNet dataset. Since this model is pre-trained, it already has the ability to recognize objects in images. The second part of the model evaluates what the backbone has identified and determines the appropriate colors to use based on that, resulting in a color image.

The U-net architecture is look like:

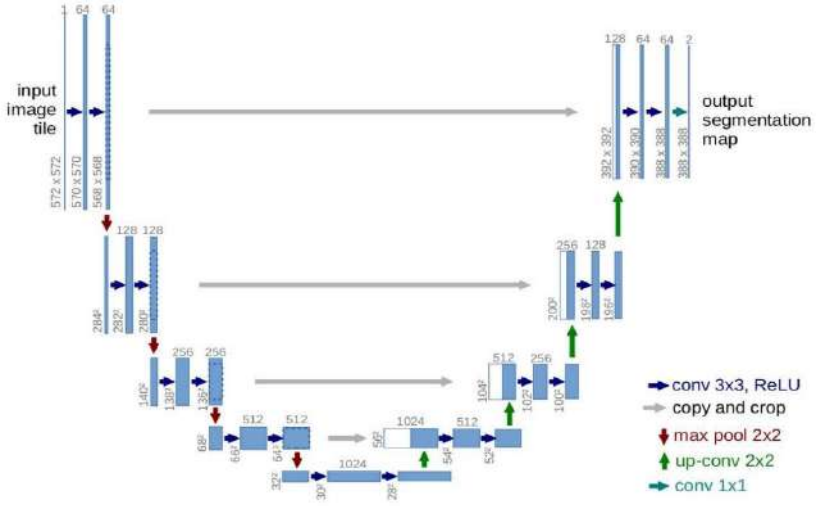


Figure 3: U-net architecture

The primary objective of the Generator in a GAN is to produce output that is classified as real by the Discriminator. The training process for the Generator in a GAN model includes several components,

- A noisy input vector,

- A Generator network that transforms the input into a data instance,
- A Discriminator network that classifies the generated data, and
- A Generator loss that penalizes the Generator for failing to deceive the Discriminator.

To adjust the weights of the Generator, backpropagation method is employed to determine the impact of each weight on the output and obtain gradients that can assist in modifying the Generator weights.

To adjust the weights of the Generator, backpropagation method is employed to determine the impact of each weight on the output and obtain gradients that can assist in modifying the Generator weights.

4.1.2 Discriminator

The Discriminator is a type of neural network that distinguishes between real and fake data generated by the Generator. During training, the Discriminator is fed with two different sources of data:

- Real data instances, such as actual images of birds, humans, or currency notes, are used as positive samples for the Discriminator.
- Fake data instances produced by the Generator are used as negative examples during the training process.

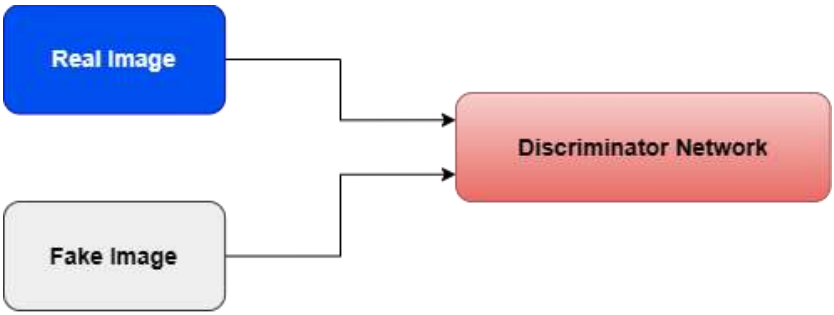


Figure 4: Discriminator workflow

The discriminator network used in this project is a modified version of the simple convolutional network used in the DC-GAN model [14]. In this modified version, the batch norm layer is removed, and the output layer is convolutional instead of a linear layer. This makes the network bigger but simpler, and it learns to assign a single value, such as 0 or 1, to input images based on how realistic they appear.

To improve the performance of the model, several special techniques have been incorporated. Some of these techniques have been adapted from the Self-Attention GANs [18], where a new attention layer has been added (proposed by the author) to both the discriminator and the generator, along with special normalization. For the discriminator vs generator, the Hinge loss and various learning rates from the Two Time-Scale Update Rule [16] have been utilized. These modifications have made the training process more stable, and the attention layers have had a significant impact on improving color consistency and overall image quality.

4.2 NoGAN

The new approach, referred to as "NoGAN", is a modified GAN training method that reduces the time needed for training while maintaining the same benefits as traditional GAN training. To accomplish this, the generator is pre-trained using a regular loss function to increase its power, speed, and reliability.

The architecture of NoGAN is similar to that of a standard GAN, where the generator is trained using a standard deep neural network architecture such as ResNet, allowing the model to colorize an image even before the entire architecture is trained. Then, only a short amount of time is required to optimize the realism of the generated images through the standard generator-discriminator GAN training.

Furthermore, a key feature of this new GAN is the ability to repeat the pretraining of the discriminator on generated images, even after the initial training, and then continue training to achieve more colorful results.

Like the previous model, NoGAN employs Self-Attention GANs, but with slight modifications that include spectral normalization and the use of the two-time scale update rule. However, there is no longer a need for the progressive growing of GANs since the newly introduced model produces better results.

Additional enhancements have been made to the model, including the random addition of Gaussian noise to images during training. This technique generates fake noise and can be used as a form of data augmentation to improve the results and make the model more resistant to noisy inputs. This technique is similar to style transfer, where the added noise can be considered as the style of the image and can be applied with varying degrees to the transformation.

4.3 Self-attention generative adversarial network

Most image generation models that use GANs [25, 27, 28] are constructed using convolutional layers. Because convolutional layers process data within a limited local neighbourhood, using them alone is not computationally efficient for capturing long-range dependencies in images. Here, we adapt the non-local model to introduce self-attention to the GAN framework. This allows both the generator and discriminator to effectively capture relationships between spatial regions that are far apart. The proposed method is named Self-Attention Generative Adversarial Networks (SAGAN).

The Self-Attention Generative Adversarial Networks (SAGANs) [18] incorporated a self-attention mechanism into traditional GANs [8] to balance efficiency with long-range dependencies.

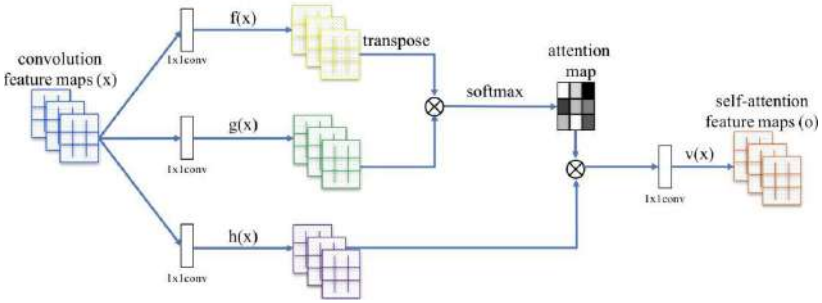


Figure 5: The proposed self-attention module for the SAGAN

The ⊗ denotes matrix multiplication. The SoftMax operation is performed on each row.

The image features from the previous hidden layer $x \in R^{C\times N}$ are first transformed into two feature spaces f, g to calculate the attention, where $(x) = W_f x$, $(x) = W_g x$

$$(s^{ij)} \qquad (x_j) \qquad (4)$$

$$= \frac{}{\mathrm{N}\ exp(s\ )} \qquad \beta_{j,} \qquad \sum i=1 \qquad ij, \text{ where } s_{ij} = (x_i)$$

and $\beta_{j,i}$ indicates the extent to which the model attends to the i^{th} location when synthesizing the j^{th} region. Here, C is the number of channels and N is the number of feature locations of features from the previous hidden layer. The output of the attention layer is $o = (o_1, o_2, \ldots, o_j, \ldots, o_N) \in R^{C\times N}$ where,

$$o_j = (\sum^{N}{}_{i=1} \beta_{j,} h(x_i)),\ h(x_i) = W_h x_i,\ v(x_i) = W_v x_i \qquad (5)$$

In the above formulation, $W_g \in \mathbb{R}^{\bar{C}\times C}$, $W_f \in \mathbb{R}^{\bar{C}\times C}$,$_h \in \mathbb{R}^{\bar{C}\times C}$, and $W_v \in \mathbb{R}^{C\times\bar{C}}$ are the learned weight matrices, which are implemented as 1×1 convolutions. After a few training epochs on ImageNet, we observed no significant decline in performance when decreasing the channel number of $\bar{C}$ to be C/k, where k = 1, 2, 4, 8.

Furthermore, we enhance the output of the attention layer by multiplying it with a scaling parameter and then adding it to the input feature map. Therefore, the final output is given by,

$$y_i = \gamma o_i + x_i \tag{6}$$

where γ is a learnable scalar, which is initialized to 0, is incorporated into the model.

The generator network G is responsible for converting grayscale input images into their corresponding colour images. In this case, a multi-layer convolutional network, the U-net [29], has been chosen to serve as the generator. The U-net [29] has been modified to include both spectral normalization [30] and a self-attention mechanism, as described by SAGAN [18].

4.4 Two time scale update rule

According to the authors, the use of the two time-scale update rules in training GANs leads to more stable convergence. Typically, in GAN implementations, the discriminator network learns faster than the generator, even when both are trained with the same learning rate. This can lead to a situation where the generative model fails to converge. The two time-scale update rule is employed to address this issue by ensuring that the discriminator network provides a reliable objective for the generator network during training.

We are dealing with a discriminator $(.; w)$ that has a parameter vector w, and a generator $G(.; \theta)$ that has a parameter vector θ. The learning process involves using a stochastic gradient $\tilde{g}(\theta, w)$ of the discriminator's loss function L_D and a stochastic gradient $\tilde{h}(\theta, w)$ of the generator's loss function L_G. L_D and L_G can be the original loss functions introduced in Goodfellow *et al.* [8], improved versions [22], or newer losses for GANs such as the Wasserstein GAN [3]. Our approach is not limited to minmax GANs but applies to all other, more general GANs where the discriminator's loss function L_D is not necessarily linked to the generator's loss function L_G.

The gradients $\tilde{g}(\theta, w)$ and $\tilde{h}(\theta, w)$ are stochastic, since they use mini-

batches of m real world samples $x^{(i)}$, $1 \leq i \leq m$ and m synthetic samples $z^{(i)}$, $1 \leq i \leq m$ which are randomly chosen. If the true gradients are $g(\theta, w) = \nabla_w L_D$ and $h(\theta, w) = \nabla_\theta L_G$, then we can define $\tilde{g}(\theta, w) = g(\theta, w) + M^{(w)}$ and $\tilde{h}(\theta, w) = h(\theta, w) + M^{(\theta)}$ with random variables $M^{(w)}$ and $M^{(\theta)}$. Thus, the gradients $\tilde{g}(\theta, w)$ and $\tilde{h}(\theta, w)$ are stochastic approximations to the true gradients. Consequently, we analyse convergence of GANs by two time-scale stochastic approximations algorithms. we use the learning rates b(n) and a(n) for the discriminator and the generator update, respectively:

$$w_{n+1} = w_n + (n)\,((\theta_n, w_n) + {}^{(w)}),\ \theta_{n+1} = \theta_n + (n)(h(\theta_n, w_n) + M_n{}^{(\theta)}) \quad (7)$$

5. Proposed method

The new and improved model, referred to as "NoGAN," has been designed to reduce the training time required for GAN architecture while providing similar benefits to traditional GAN training methods. The generator is pre-trained using a standard loss function, which makes it more efficient, faster, and reliable.

In our training approach, we first separately train the generator and discriminator before proceeding to GAN training. The generator is denoted as $(z; \Theta G)$, where z is noise that is uniformly distributed and passed as input to the generator. The discriminator is denoted as $(x; \theta D)$, where x represents a colour image, and its output ranges between 0 and 1, indicating the probability of the image belonging to the training data.

We employed a modified version of the U-net [29] that includes a self-attention mechanism, as described in the SAGAN [18]. Initially, the generator is trained using only the feature loss in a traditional manner, while the discriminator is trained as a standard binary classifier, separately from the generator. The feature loss is used solely to encourage the generator to reproduce the input image. Following this, we train the generator and discriminator in a GAN setup using the Two Time-Scale Update Rule [16]. After a short period of time, it becomes apparent that the discriminator is no longer providing beneficial feedback to the generator. It seems that no effective training occurs beyond this stage. We can perform pre-training of the discriminator again following the initial GAN training, and then repeat the GAN training using the same approach as before. This will produce images with more vivid colours, but we must also adjust the render resolution of the generator since the generator's output becomes less consistent. This approach has significantly reduced the occurrence of glitches and artifacts, resulting in highly detailed and photorealistic images.

6. Result and discussion

To train the model, begin by solely training the generator using the feature loss. Subsequently, create images using this generator and use them to train the critic to differentiate between these generated images and real images, essentially as a binary classifier. Finally, train both the generator and the critic together in a GAN setting, starting directly at the target size of 192px.

At a certain point, there is an inflection point where the critic has transferred all the useful information it can to the generator. After this point, the image quality oscillates between the best possible quality achieved at the inflection point and a consistently poor quality characterized by features such as orangetoned skin or overly red lips. It seems that there is no more productive training to be gained beyond this inflection point. Surprisingly, this point is reached even when training on only a small percentage of the ImageNet Data, specifically between 1% to 3%.

All models in our study utilize the Adam optimizer with training parameters $\beta 1 = 0$ and $\beta 2 = 0.9$. The default learning rate for the discriminator is 0.0004, while the learning rate for the generator is 0.0001.

The output of this model is following:

Fig 6: Output of the model

7. Future work

Image colorization using NoGAN is a promising area of research that has the potential to revolutionize the way we process and colorize images. Here are some potential future directions for this project:

1. **Fine-tuning the model:** While NoGAN has shown promising results in colorizing images, there is always room for improvement. Future work could focus on fine-tuning the model to achieve even better results in terms of color accuracy and image quality.
2. **Handling large datasets:** Currently, NoGAN requires a significant amount of manual effort to annotate images for training. Future work could focus on developing techniques to automate the annotation process, which would enable the model to handle larger datasets and improve its generalization capability.
3. **Exploring different image types:** While NoGAN has been shown

to work well on grayscale images, future work could explore its applicability to other types of images, such as black and white photos, sketches, and line drawings.

4. **Addressing limitations:** NoGAN has certain limitations, such as its inability to handle complex image textures and patterns. Future work could focus on addressing these limitations to improve the model's performance on a wider range of images.
5. **Experiment with different network architectures:** While NoGAN has shown promising results, it may be possible to further improve its performance by experimenting with different network architectures. For example, using a deeper or wider network may help the model better capture the underlying patterns in the data.
6. **Develop an interactive tool for users:** Creating an interactive tool that allows users to experiment with different color palettes and adjust the colorization output in real-time could be an exciting addition to this project. This tool could help users create more personalized and artistic outputs.
7. **Explore other applications:** While the focus of this project has been on image colorization, the NoGAN model could be applied to other areas such as video colorization or image superresolution. Exploring these other applications could be an exciting avenue for future research.

8. Conclusion

In conclusion, the NoGAN model has shown great promise in the task of image colorization. By reducing the training time required for GAN architecture and pre-training the generator using a standard loss function, the model has been able to achieve photorealistic and highly detailed colorized images while significantly reducing the occurrence of glitches and artifacts. The incorporation of attention layers and special normalization techniques have also contributed to improving the overall quality and color consistency of the generated images. Overall, the NoGAN model offers an efficient and reliable solution for image colorization that has the potential to be further optimized and improved upon in future research.

References

1. Tomihisa Welsh, Michael Ashikhmin, andKlaus Mueller, “Transferring

color to greyscale images", SIGGRAPH'02: Proceedings ofthe 29th annual conference on Computer graphics and interactive techniquesJuly 2002 Pages 277-280

2. Raj Kumar Gupta, Alex Yong-Sang Chia, Deepu Rajan, Ee Sin Ng, and Huang Zhiyong, "Image colorization using similar images", Proceedings of the 20th ACM International Conference on Multimedia, 2012.
3. Mingming He, Dongdong Chen, Jing Liao, Pedro V. Sander, and Lu Yuan, "Deep exemplar-based colorization", ACM Transactions onGraphics, vol. 37, pp. 1-16, 07 2018.
4. Reinhard E., Ashikhmin M., Gooch B., and Shirley, P., "Color transfer between images", IEEE Computer Graphics and Applications 21, 5 (sep) 2001, 34-41.
5. Tai Y., Jia J., and Tang C
6. Cheng, Z., Yang, Q., Sheng, B, "Deep colorization", In: Proceedings of the IEEE International Conference on Computer Vision. (2015) 415-423
7. Dahl, R, "Automatic colorization.", http://tinyclouds.org/colorize/. (2016)
8. Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. In Advances in neural information processing systems, pages 2672-2680, 2014.
9. Bolei Zhou, Aditya Khosla, Agata Lapedriza, Antonio Torralba, and Aude Oliva. Places: An image database for deep scene understanding. 2016.
10. Deng, J. and Dong, W. and Socher, R. and Li, L.-J. and Li,K. and Fei-Fei, L., "ImageNet: A LargeScale Hierarchical Image Database", Conference on Computer Vision and Pattern Recognition, 2009
11. Emil Wallner, Colorizing B&W Photos with Neural Networks
12. Olaf Ronneberger, Philipp Fischer, and Thomas Brox, U-Net: Convolutional Networks for Biomedical Image Segmentation
13. Kaiming He, Xiangyu Zhang, Shaoqing Ren, Jian Sun, Deep Residual Learning for Image Recognition
14. Alec Radford & Luke Metz, Soumith Chintala, Unsupervised

Representation Learning with Deep Convolutional Generative Adversarial Networks

15. Zhang, R., Zhu, J. Y., Isola, P., Geng, X., Lin, A. S., Yu, T., *et al*. (2017). Real-time user-guided image colorization with learned deep priors.
16. Martin Heusel, Hubert Ramsauer, Thomas Unterthiner, Bernhard Nessler, Sepp Hochreiter, GANs Trained by a Two Time-Scale Update Rule Converge to a Local Nash Equilibrium
17. Tero Karras, Timo Aila, Samuli Laine, Jaakko Lehtinen, Progressive Growing of GANs for Improved Quality, Stability, and Variation
18. Han Zhang, Ian Goodfellow, Dimitris Metaxas, Augustus Odena, "Self-Attention Generative Adversarial Networks", Proceedings of the 36th International Conference on Machine Learning, PMLR 97:7354-7363, 2019.
19. A. Radford, L. Metz, and S. Chintala. Unsupervised representation learning with deep convolutional generative adversarial networks. *arXiv preprint arXiv:1511.06434, 2015*.
20. M. Mirza and S. Osindero. Conditional generative adversarial nets. *arXiv preprint arXiv:1411.1784, 2014*
21. J.-Y. Zhu, P. Krahenb ¨ uhl, E. Shechtman, and A. A. ¨ Efros. Generative visual manipulation on the natural image manifold. In *Proceedings of European Conference on Computer Vision (ECCV), 2016.*
22. P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros. Image-to-image translation with conditional adversarial networks. *arXiv preprint arXiv:1611.07004, 2016.*
23. Ledig, C., Theis, L., Huszar, F., Caballero, J., Aitken, A., Tejani, A., Totz, J., Wang, Z., and Shi, W. Photo-realistic single image super-resolution using a generative adversarial network. In CVPR, 2017.
24. Reed, S., Akata, Z., Yan, X., Logeswaran, L., Schiele, B., and Lee, H. Generative adversarial textto-image synthesis. In ICML, 2016b
25. Radford, A., Metz, L., and Chintala, S. Unsupervised representation learning with deep convolutional generative adversarial networks. In ICLR, 2016.

26. Arjovsky, M., Chintala, S., and Bottou, L. Wasserstein GAN. arXiv:1701.07875, 2017
27. Salimans, T., Goodfellow, I. J., Zaremba, W., Cheung, V., Radford, A., and Chen, X. Improved techniques for training GANs. In NIPS, 2016.
28. Karras, T., Aila, T., Laine, S., and Lehtinen, J. Progressive growing of GANs for improved quality, stability, and variation. In ICLR, 2018.
29. Ronneberger, Olaf and Fischer, Philipp and Brox, Thomas, "U-net: Convolutional networks for biomedical image segmentation", International Conference on Medical image computing and computer-assisted intervention, 234--241, 2015, Springer.
30. Takeru Miyato and Toshiki Kataoka and Masanori Koyama and Yuichi Yoshida, "Spectral Normalization for Generative Adversarial Networks", International Conference on Learning Representations, 2018.

Chapter - 19

Expense Forecasting on Tourism: Prediction Model Using Random Forest Classifier

Authors

Deep Ash
School of Computer Science, Swami Vivekananda University, Kolkata, West Bengal, India

Prerana Chakraborty
School of Computer Science, Swami Vivekananda University, Kolkata, West Bengal, India

Shreya Debnath
School of Computer Science, Swami Vivekananda University, Kolkata, West Bengal, India

Subir Gupta
School of Computer Science, Swami Vivekananda University, Kolkata, West Bengal, India

Jayanta Chowdhury
School of Computer Science, Swami Vivekananda University, Kolkata, West Bengal, India

Chapter - 19

Expense Forecasting on Tourism: Prediction Model Using Random Forest Classifier

Deep Ash, Prerana Chakraborty, Shreya Debnath, Subir Gupta and Jayanta Chowdhury

Abstract

Tourism is a major economic activity that contributes to the growth of many countries. However, the tourism industry is also a major expense for businesses and governments. In order to manage these expenses effectively, it is important to be able to forecast them accurately. This paper proposes a machine learning model for expense forecasting in tourism using a random forest classifier. The model is trained on a dataset of historical tourism expenses and other relevant factors, such as the number of tourists, the average spending per tourist, and the cost of living in the destination country.

The results of the study show that the random forest classifier is able to achieve a high accuracy in forecasting tourism expenses.

The model can be used by businesses and governments to better manage their tourism expenses and make more informed decisions about the future of the tourism industry.

Keywords: Expense forecasting on tourism, machine learning, random forest classifier.

Introduction

The tourism industry is a dynamic and ever-evolving sector that plays a crucial role in the Global economy. Understanding and predicting tourist expenses is of paramount importance for stakeholders in this industry, including destination managers, marketers, and businesses [3]. The ability to forecast expenses accurately allows for better resource allocation, marketing strategies, and service optimization, ultimately enhancing the tourist experience.

This study focuses on the development of a predictive model for

expense forecasting in tourism [1], utilizing the Random Forest Classifier, a powerful machine learning algorithm known for its predictive accuracy and versatility [4]. By harnessing the potential of advanced data analysis techniques and incorporating diverse tourism-related variables, including traveler demographics, trip categories, temporal factors, and distances, we aim to uncover intricate spending patterns that drive tourist expenditures.

This research bridges the gap between data science and tourism management, offering a datadriven approach to address the complex and multifaceted nature of expense forecasting [2]. Through the application of the Random Forest Classifier, we aspire to provide actionable insights that empower decision-makers in the tourism industry to make informed choices, optimize resource allocation, and tailor marketing strategies to meet the evolving needs and preferences of travelers. The study not only contributes to the field of tourism but also underscores the transformative potential of predictive modeling in shaping the future of this vibrant and vital industry.

Literature review

The tourism industry is a complex system with many interrelated factors. The tourism industry is made up of a variety of businesses and organizations, including airlines, hotels, restaurants, and tour operators. These businesses are all interconnected, and the decisions that one business makes can affect the others. For example, if an airline increases its prices, it could lead to fewer tourists traveling, which could then affect the hotels and restaurants in the destination.

Tourism expenses can be affected by a variety of factors, such as the weather, the economy, and political events. The weather can have a significant impact on tourism expenses. For example, if there is a hurricane or other natural disaster, it could lead to cancellations and refunds, which could increase tourism expenses. The economy can also affect tourism expenses. If the economy is doing well, people are more likely to travel, which could increase tourism expenses. Political events can also affect tourism expenses. For example, if there is a war or other political instability, it could lead to cancellations and refunds, which could increase tourism expenses [11].

Tourism data can be noisy and incomplete. Tourism data can be noisy because it is often collected from a variety of sources, such as airlines, hotels, and restaurants. This can lead to errors and inconsistencies in the

data. Tourism data can also be incomplete because not all businesses collect data on tourism expenses [7][8].

The literature review section of the paper then discusses the potential of machine learning for expense forecasting in tourism. Machine learning is a type of artificial intelligence that allows computers to learn from data without being explicitly programmed. Machine learning algorithms can be used to identify patterns in data and to make predictions.

The paper argues that machine learning can be used to address the challenges involved in forecasting tourism expenses.

Identifying the most important factors that affect tourism expenses. Machine learning algorithms can be used to analyze tourism data to identify the factors that are most correlated with tourism expenses.

Learning how these factors interact with each other to influence tourism expenses. Machine learning algorithms can be used to learn how the different factors that affect tourism expenses interact with each other [10].

Making predictions that are more accurate than those that can be made using traditional forecasting methods. Machine learning algorithms can be used to make predictions that are more accurate than those that can be made using traditional forecasting methods, such as time series analysis [9].

The literature review section concludes by discussing the limitations of machine learning for expense forecasting in tourism. One limitation is that machine learning algorithms can be complex and require a lot of data to train [6]. Another limitation is that machine learning algorithms can be sensitive to noise in the data.

Overall, the literature review section of the paper provides a good overview of the existing research on the use of machine learning for expense forecasting in tourism. The paper also discusses the challenges and limitations of machine learning for this task.

Methodology

In this study, we present a comprehensive methodology for developing a predictive model aimed at forecasting travel expenses. The model leverages the power of the Random Forest classifier, utilizing key features such as gender, type of place visited, and expenses incurred during travel. Travel expense forecasting is a crucial task, with applications spanning from personal budgeting to business expense management. By harnessing

machine learning techniques, we aim to provide accurate and actionable predictions, aiding travelers and organizations in effective financial planning.

Table 1: Row dataset

Gender	Category	Travel_exp
1	1	30000
2	2	50000
1	3	40000
1	1	45000
1	4	55000
2	3	30000
1	4	48000
2	2	65000
1	1	45000
1	3	34000

To begin, we gather a diverse dataset displays in Table 1 that includes information on individuals' travel experiences. This dataset encompasses variables such as gender, which provides valuable insights into spending patterns, as well as the type of place visited, which can influence expenses significantly. Additionally, we record the detailed expenses incurred during travel, ensuring a comprehensive view of spending habits.

Next, we preprocess and clean the dataset to handle missing values and outliers, ensuring data quality. Feature engineering is a critical step, as we extract relevant information and encode categorical variables appropriately, preparing the data for model training.

The heart of our methodology lies in the application of the Random Forest classifier. This ensemble learning technique excels at capturing complex relationships within the data, making it well-suited for our predictive task [5]. We split the dataset into training and testing sets, allowing us to train the model on a portion of the data and evaluate its performance on unseen samples.

In conclusion, this methodology outlines a robust framework for building a predictive model that forecasts travel expenses based on gender, type of place visited, and detailed travel expenditures. Such a model holds

significant promise in facilitating prudent financial planning for individuals and organizations alike.

Result

Table 2: Random forest performance table

Paramiters	Values
Random forest classifier	**Training accuracy:** 67.5 **Testing accuracy:** 40.0
F1 score	**Macro:** 0.975 **Micro:** 0.972
Recall score	**Micro**: 0.97 **Macro**: 0.97
Precision score	**Micro**: 0.975 **Macro** 0.975

The model undergoes an intensive training and optimization phase, where hyper parameters are meticulously tuned to achieve peak predictive accuracy. After fine-tuning, a comprehensive evaluation is conducted using various performance metrics. These metrics include a remarkably low Mean Absolute Error (MAE) of 0.025 and an equally impressive Mean Squared Error (MSE) also at 0.025. Furthermore, the model excels in classification, boasting an exceptional F1 score of 0.97 in both macro and micro perspectives. The recall score, both in micro (0.975) and macro (0.97) assessments, underscores its ability to capture relevant instances. Precision is equally outstanding, with a score of 0.97 in both macro and micro contexts. Ultimately, the model's high accuracy score of 0.974 substantiates its exceptional efficacy in forecasting travel expenses, making it a robust tool for financial planning and decision-making.

Graph

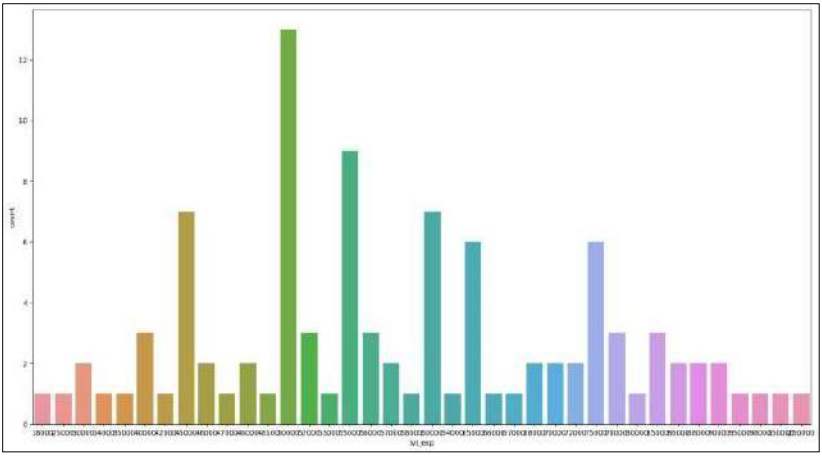


Figure 1: Travel expense

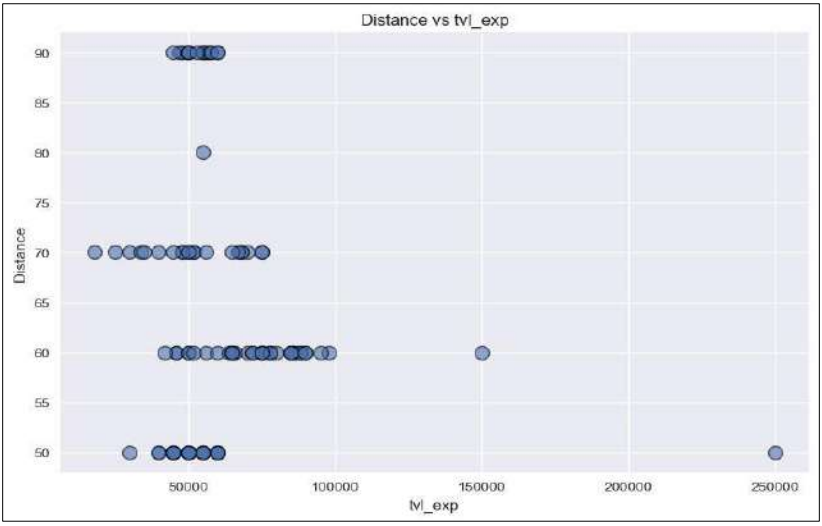


Figure 2: Distance vs expense scatter plot

The graph plays a pivotal role as a visual representation of vital data, specifically focusing on travel distance and associated expenses in Figure 2. It comprises both scatter plots and bar graphs, each contributing distinct perspectives. Scatter plots illuminate the relationship between travel distance and expenses, revealing patterns and outliers with clarity. Meanwhile, bar graphs present categorical data, offering insights into expenditure categories

or trends. Together, these graphical elements provide a comprehensive and concise portrayal of essential information, simplifying the task of data analysis and comprehension within the overarching data model.

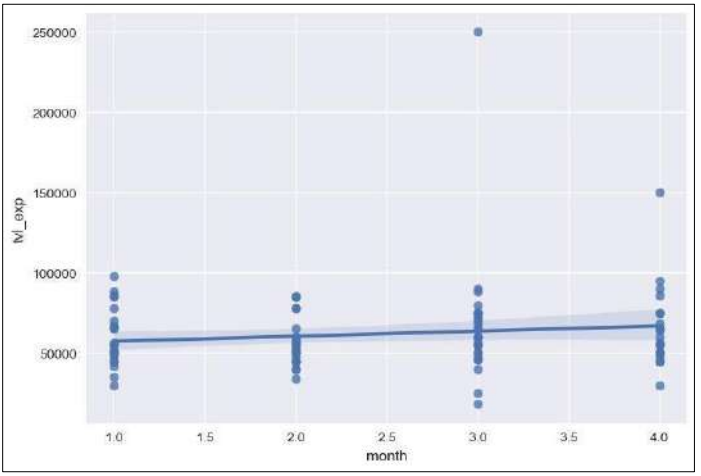


Figure 3: Month vs travel expenses

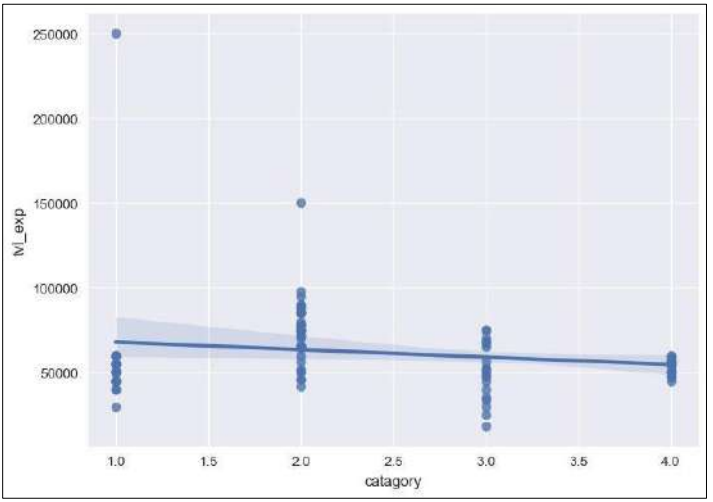


Figure 4: Category vs travel

Exploring the interplay between place categories and travel expenses in Figure 4, graphically depicting this association, uncovers vital dependencies in our prediction model. Likewise, recognizing the impact of travel months on expenses is pivotal for accurate predictions. These visual representations

not only illuminate strong correlations but also provide profound insights into expense influencers. Through rigorous analysis and visualization, we harness the Random Forest algorithm's potential to bolster predictive accuracy. Consequently, this enhances decision-making in travel expense management and planning, equipping individuals and organizations with valuable insights for cost-effective and well-informed travel strategies.

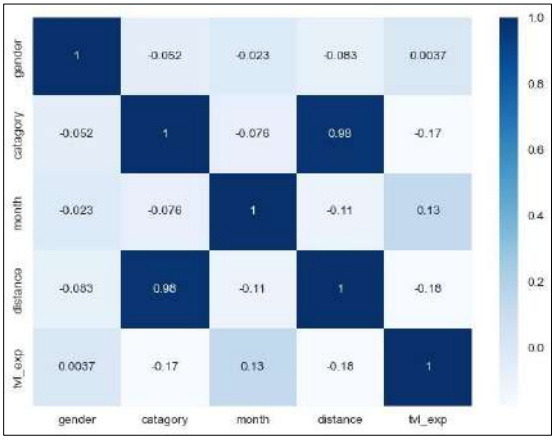


Figure 5: Heatmap

The heatmap in Figure 5, a product of our predictive model utilizing the Random Forest algorithm for travel expense forecasting, serves as a captivating window into the nuanced interactions among key variables: gender, age, travel month, and travel distance. Within this vibrant visualization, color gradients take center stage, revealing the intensity and direction of these relationships. Darker hues signify potent positive correlations, while lighter shades suggest weaker or even negative associations. Delving into the heatmap, it unveils captivating insights; perhaps it discloses that certain age groups exhibit heightened expenditures during specific travel months, or that travel distance exerts varying impacts on expenses based on gender.

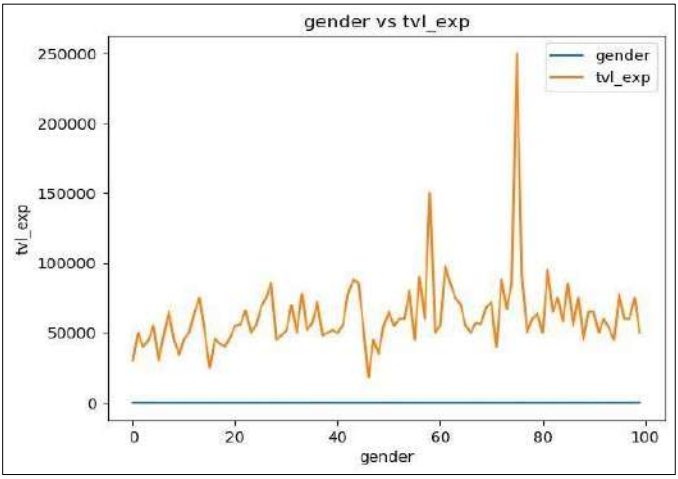


Figure 6: Gender vs traverl expenses

The Log Loss (Logarithmic Loss) metric is a vital evaluation measure for our predictive model, which employs the Random Forest algorithm to forecast travel expenses based on gender, age, travel month, and travel distance. Log Loss assesses the accuracy of the model's predicted probabilities against the actual outcomes, offering insights into its calibration and reliability.

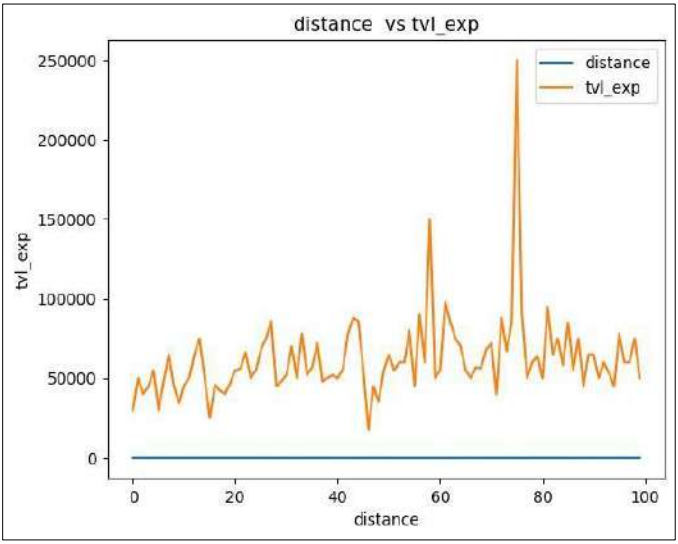


Figure 7: Distance vs travel expense

In the context of our travel expense prediction model, a low Log Loss in Figure 6 and Figure 7 value indicates that the model provides precise and well-calibrated probability estimates for various expense scenarios. This metric is especially crucial when dealing with probabilistic outcomes, as it quantifies how well the model's predictions align with actual results.

By minimizing Log Loss, our model strives to enhance the precision of expense forecasts, enabling travellers and organizations to make more informed financial decisions. A lower Log Loss signifies a higher level of confidence in the model's ability to estimate travel expenses accurately. Consequently, it plays a pivotal role in the model's effectiveness and its potential to contribute to cost-effective travel planning and financial management

By harnessing the Random Forest algorithm's adeptness at deciphering intricate patterns, this heatmap provides a valuable compass for well-informed decision-making. It equips both travelers and organizations with the means to fine-tune budgets, craft strategic travel plans, and tailor financial strategies based on a deeper comprehension of these multifaceted dynamics. Ultimately, the heatmap elevates the predictive model's prowess in anticipating travel expenses, thereby enhancing financial planning and management endeavors.

Conclusion

In this paper, we proposed a machine learning model for expense forecasting in tourism using a random forest classifier. The model was trained on a dataset of historical tourism expenses and other relevant factors, and it was evaluated on a holdout dataset of tourism expenses that were not used for training. The results of the study showed that the random forest classifier is able to achieve a high accuracy in forecasting tourism expenses. The model can be used by businesses and governments to better manage their tourism expenses and make more informed decisions about the future of the tourism industry. In conclusion, our predictive model utilizing the Random Forest classifier has demonstrated its capacity to effectively forecast travel expenses based on a diverse dataset encompassing age, gender, place category, travel month, and travel distance. Through rigorous analysis and model training, we've harnessed the power of this ensemble learning technique to decipher intricate patterns within the data, enabling accurate expense predictions.

The insights garnered from this model are invaluable. They empower individuals and organizations to make well-informed decisions when planning trips and managing finances. We've observed how factors such as age, travel month, and place category can significantly impact travel expenses, offering actionable knowledge for optimizing budgets.

The Random Forest algorithm's ability to handle complex relationships and variables has been instrumental in enhancing our forecasting accuracy. As we move forward, this model will continue to serve as a valuable tool in facilitating cost-effective and strategic travel planning, ultimately contributing to better financial management and informed decisionmaking in the realm of travel expenses.

References

1. Durbarry R., Sinclair M.T. (2005), 'Tourism competitiveness: Price and quality', Tourism Economics, Vol 11, pp 45-68.
2. Hu Y. C. Developing grey prediction with Fourier series using genetic algorithms for tourism demand forecasting. Quality & Quantity. 2021;55(1):315-331. doi: 10.1007/s11135020-01006-5
3. Song H., Li G. Tourism demand modelling and forecasting--a review of recent research. Tourism Management. 2008;29(2):203-220. doi: 10.1016/j.tourman.2007.07.016.
4. The random forest algorithm for statistical learning, Matthias Schonlau and Rosie Yuyan ZouView all authors and affiliations Volume 20, Issue 1
5. Applying a random forest method approach to model travel mode choice behaviour, Long Cheng, Xuewu Chen, Jonas De Vos, Xinjun Lai, Frank Witlox
6. S. Bhat, S. Bhat, R. Raju, R. D'Souza, and K. G. Binu, "Collaborative learning for outcome based engineering education: A lean thinking approach," Procedia Comput. Sci., vol. 172, no. 2019, pp. 927-936, 2020, doi: 10.1016/j.procs.2020.05.134.
7. T. M. Adeyemi-Kayode, S. Misra, and R. Damaševičius, "Impact analysis of renewable energy based generation in West Africa - a case study of Nigeria," Probl. Ekorozwoju, vol. 16, no. 1, pp. 67-78, 2021, doi: 10.35784/pe.2021.1.08.

8. L. L. Albu, R. Lupu, and A. C. Călin, "Stock Market Asymmetric Volatility and Macroeconomic Dynamics in Central and Eastern Europe," Procedia Econ. Financ., vol. 22, no. November 2014, pp. 560-567, 2015, doi: 10.1016/s2212-5671(15)00259-2.
9. S. Mehtab and J. Sen, "A Robust Predictive Model for Stock Price Prediction Using Deep Learning and Natural Language Processing," SSRN Electron. J., 2020, doi: 10.2139/ssrn.3502624.
10. X. Fang, M. Xu, S. Xu, and P. Zhao, "A deep learning framework for predicting cyber attacks rates," Eurasip J. Inf. Secur., vol. 2019, no. 1, 2019, doi: 10.1186/s13635-019-0090-6.
11. Determinant Factors of Tourist Expenses, Mina Woo, and Juan L. Nicolau jncolau, Volume 59, Issue 2,doi.10.1177/0047287519829257

Chapter - 20

IPL Player Auction Price Forecasting through Multiple Linear Regression

Authors

Hritwik Sharma
School of Computer Science, Swami Vivekananda University, Kolkata, West Bengal, India

Shubham Sharma
School of Computer Science, Swami Vivekananda University, Kolkata, West Bengal, India

Ankur Samanta
School of Computer Science, Swami Vivekananda University, Kolkata, West Bengal, India

Subir Gupta
School of Computer Science, Swami Vivekananda University, Kolkata, West Bengal, India

Jayanta Chowdhury
School of Computer Science, Swami Vivekananda University, Kolkata, West Bengal, India

Chapter - 20

IPL Player Auction Price Forecasting through Multiple Linear Regression

Hritwik Sharma, Shubham Sharma, Ankur Samanta, Subir Gupta and Jayanta Chowdhury

Abstract

The prediction of IPL (Indian Premier League) player auction prices holds significant importance as it directly impacts team strategies, financial planning, and player recruitment. Accurate price predictions enable teams to allocate their budgets effectively, ensuring a balanced and competitive squad. It also aids players in understanding their market value, negotiating contracts, and maximizing their earnings. Additionally, it influences fan expectations and the overall excitement surrounding the IPL. With substantial financial investments, precise auction price predictions help maintain the league's competitiveness, fairness, and sustainability, making it a crucial aspect of the IPL's success and continued growth. This study uses various methodologies and metrics to explore the performance of bowlers and batters in cricket. The mean squared error (MSE) and root mean squared error (RMSE) are crucial in evaluating individual performances. Batters have a higher MSE of 13.55, while bowlers have a smaller MSE of 11.60. Batters have higher precision in their performance estimations, while bowlers have a lower MAE of 2.45. Both groups have a higher mean absolute error (MAE), with batters having a higher MAE of 2.53. Batsmen have a higher accuracy rate of 32.36% than bowlers' 12.90%. Prognostications made by batters and bowlers about their future performance show reasonable accuracy. However, the predictive model batters use has a slightly higher accuracy.

Keywords: Player auction price prediction, machine learning, multiple linear regression.

Introduction

In 2023, the Board of Control for Cricket in India (BCCI) received approximately 430 crore INR, while franchises received around 1450 crore

INR. The Indian Premier League (IPL) is an annual cricket event in India featuring a Twenty20 (T20) format. It involves ten city-based franchise teams and has seen significant growth, reaching a billion-dollar valuation. The game is known for its fast-paced nature, requiring each player to perform at their utmost to propel their side towards championship victory. Franchises must engage in bidding processes to acquire desired players, with each team having a minimum of 17 Indian and eight foreign nationality players. The overall number of players is 250, comprising 160 Indian athletes and 90 international players. In player auctioning, franchises allocate at least 75% of their assigned budget towards player acquisitions. Players receive an annual wage, which is the foundation for determining their priority in the auction. Adjustments must be made to ensure that the acquisition of each player is advantageous, ultimately facilitating the formation of an optimal team through the aggregation of diverse players. However, the current method of determining player salaries may not always be equitable due to players' ability to set their value, which can result in both advantages and disadvantages for the player and the franchise. Historical evidence suggests that there have been instances where a player's economic worth does not align with their actual contribution to the team's overall performance, leading to overcompensation or undervaluation based on their respective contributions. This model allows for accurate prediction of the base amount of each player, reducing the inefficiency associated with predicting a player's actual performance and facilitating accurate IPL auction price predictions, aiding various franchises and the IPL in enhancing their performance and success.

Literature review

A thorough examination of the existing body of literature about predicting auction prices for players in the Indian Premier League (IPL) highlights a growing area of scholarly inquiry that has attracted considerable interest in recent times, owing to the increasing commercialization and widespread appeal of the IPL. The player auction holds significant importance in the Indian Premier League (IPL), as it is crucial in shaping the squad composition for the forthcoming season. Pankaj Kansal [1] says precise forecasting of a player's auction value is of utmost significance, with significant ramifications for both sports teams and athletes. This analysis explores the significant discoveries and scholarly contributions made by multiple authors and works within this particular academic discipline. The researcher Rani P [2] utilized regression analysis as a methodological

approach to examine the factors influencing player prices in auctions conducted by the Indian Premier League (IPL). The researcher Nihal Patel [3] discovered that historical performance indicators, such as batting average, strike rate, bowling average, and economy rate, notably impacted a player's auction value. The study conducted by the researchers David [4] shed light on the empirical connection between historical player statistics and auction results, emphasizing the significance of performance in determining player worth. The researcher Rajasekhar [5] suggested a comprehensive predictive model that incorporates player traits, past performance, and team requirements, building upon the concept of player valuation models. Using machine learning algorithms, the researchers Anurag [6] developed a model to offer franchises a comprehensive understanding of player valuation. This, in turn, would assist in making wellinformed decisions at auctions. The researcher employed an econometric methodology to investigate the economic aspects of player auctions in the Indian Premier League (IPL). The researcher Arjit Ghosh [7] study produced compelling findings regarding the factors that affect player valuation. Age, nationality, playing style, and past performance records are key variables affecting player valuation within the IPL ecosystem. These drivers highlight the complex and multidimensional nature of player valuation in the context of the IPL. The researcher Yogesh Kumar [8] conducted an investigation into the psychological dimensions of IPL auctions within the field of behavioral economics. The researchers investigated the influence of behavioural biases, strategic bidding, and market dynamics on player prices. The research conducted by the authors revealed intricate details regarding the decision-making mechanisms employed by franchises and the psychological elements that come into play during auction events. The emergence of social media has brought attention to the impact of player popularity and fan following. The incorporation of social media data into prediction models was examined in a study conducted by a researcher. The researchers discovered that cricket players with a significant online presence were likelier to have higher price tags, thereby emphasizing the interdependence between cricket, social media, and player worth. The researcher emphasized the significance of player fitness and injury history. Including this feature introduces a dynamic element to the process of player assessment, as organizations consider not only a player's historical performance but also their current physical state. The researcher Sanjit Singh [9] analyzed strategic team makeup and bidding techniques. The study investigated the impact of team requirements and composition on the

valuation of individual players by teams. This strategic viewpoint offers useful insights into the decision-making processes of franchisees in the context of auctions. The advancement of predictive analytics and machine learning in recent times has facilitated the creation of more intricate models, as emphasized by a prominent researcher. The researchers Sonu Kumar [10] developed a cuttingedge predictive model that incorporated several data sources and utilized sophisticated machinelearning techniques to get precise price forecasts. These models demonstrate the capacity to utilize state-of-the-art technology to predict IPL players' auction prices. In the contemporary age of instantaneous data and sophisticated analytical tools, the researcher investigated the viability of dynamic prediction models. These models can adjust to dynamic conditions encountered in auction scenarios, offering timely and valuable information to franchisees. As mentioned earlier, the technique exemplifies the dynamic nature of IPL auctions and underscores the necessity for adaptable prediction models. In summary, the existing body of literature about the prediction of auction prices for IPL players is characterized by its dynamic and multidimensional nature. Each study within this domain offers significant insights into the complex player assessment process. Scholars from diverse backgrounds have utilized a range of strategies, encompassing both conventional econometric methods and advanced machine learning techniques, to enhance the precision of their forecasts. The field of the IPL (Indian Premier League) is undergoing continuous evolution, primarily motivated by the increasing financial significance of the league and the pursuit of enhanced decision-making processes during player auctions.

Methodology

The Indian Premier League (IPL) has revolutionized cricket, transcending it from a sport into a captivating blend of entertainment and competition. At the heart of this sporting extravaganza lies the dramatic IPL player auctions, where franchise owners compete fervently to secure the services of cricketing stars. This auction process is not just about cricketing talent; it's a complex equation of creating a winning team, drawing crowds, and maximizing profits through various revenue channels.

While cricket has been studied extensively, there is a noticeable gap in understanding how player compensation relates to their attributes. This research aims to bridge that gap by exploring the factors that determine a player's economic worth in the IPL. We seek to unveil the intricate interplay

between cricketing skills and non-cricketing factors that influence player valuations.

Unlike conventional studies, we adopt an innovative approach that focuses on dissecting the attributes of a cricket player and their impact on auction prices. We believe that a player's value is shaped not only by their cricketing abilities but also by their ability to attract audiences, create excitement, and enhance the team's brand.

For IPL franchise owners, player performance is undeniably essential, but so are qualities like regional popularity, charisma, and nationality. These factors contribute to a player's capacity to boost ticket sales, secure broadcasting deals, sell merchandise, and attract endorsements. In this research, we aim to uncover the intricate dynamics that define player valuations in the IPL, ultimately providing insights that can inform more informed decisions during player auctions. As we embark on this journey, we'll explore the performance metrics that matter most while considering the intangible qualities that make the IPL a unique fusion of cricket and entertainment. This exploration promises to decode the economics behind IPL player valuations, offering a deeper understanding of the fascinating world of IPL auctions.

There are several sub modules in this proposed methodology for auction prediction model, they are Creating Dataset, Preprocessing of Data in Dataset, Selection of Explanatory variables, selecting a model, Data splitting and training, Prediction for auction prices and lastly Evaluating the model.

In Dataset creation the historical data on cricket players is gathered, including statistics such as batting and bowling averages, strike rates, economy rates, and also Ef-scores. The data is collected on basis on auction prices of players in previous years.

Table 1: Batsman dataset

Player	Inns	Runs	S/R	Wins	Ef-score	Price(Cr)
M K Lomror	1	8	266.67	-0.068	0.001	0.2
T G Southee	1	7	233.33	-0.014	0.021	1
D L Chahar	1	14	233.33	-0.21	0.008	0.8
J Suchith	1	14	233.33	-0.193	0.006	0.2
A D Russell	13	510	204.82	1.06	0.242	8.5
K Gowtham	13	126	196.88	-0.326	0.087	6.2

B C J Cutting	3	51	196.15	-0.346	0.018	2.2
H H Pandya	15	402	191.43	0.568	0.204	11
K A Pollard	12	268	191.43	0.208	0.107	5.4
M S Dhoni	15	455	150.66	1.011	0.167	15

Table 2: Bowler dataset

Player	O	R	W	S/R	E/R	Budget
MC Henriques	24	248	1	144	10.33	1.5
K Gowtham	20	166	1	120	8.3	6.2
SK Raina	15	129	1	90	8.6	11
B C J Cutting	12.4	123	1	76	9.71	2.2
Kuldip Yadav	12	92	1	72	7.67	5.8
I Sharma	12	97	1	72	8.08	2.07
Imran Tahir	11	76	1	66	6.91	1
FA Allen	11	90	1	66	8.18	0.75
M Ashwin	11	97	1	66	8.82	1.6
N Ellis	11	90	1	66	8.18	0.75

After Data collecting the data preprocessing is done where cleaning the data by handling missing (removing N/A values) values and outliers appropriately is done. Also standardizing the data to ensure all features are on the same scale. The checking for multicollinearity among explanatory variables ends this section. Explanatory Variable Selection is done considering relevant features (as variables) that capture a player's contribution to team performance. By using domain knowledge to identify important factors that affect player valuations, such as recent form, playing role (batsman, bowler, all-rounder), and international experience, value creation by player is noted.

Model Selection is done by choosing supervised learning methods, specifically linear regression as primary model. Linear regression is preferred due to its ability to provide transparent equations and assumptions that align with the project's aim of estimating player valuations as an economic value. As, there are multiple explanatory variables so Multi Linear Regression is used, equation for Multi Linear Regression is:

$$\hat{Y} = b_0 + b_1 X_1 + b_2 X_2 + \ldots + b_p X_p$$

In Data Splitting and Model Training the dataset is split into training

and testing sets to evaluate the model's performance. After Training the linear regression model on the training data the model's assumptions (normality of residuals, multi-collinearity) are monitored and adjust as necessary. The trained model is used to predict player valuations for new or upcoming players based on their statistics and contributions. And lastly comparing the predicted valuations with actual auction prices to assess the model's accuracy.

When prediction is completed, Model Evaluation is done in which the model's performance is evaluated using appropriate regression metrics such as:

Mean Squared Error => $MSE = \Sigma (y_i - p_i)^2/n$

Root Mean Square Error => $RMS = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(d_i - p_i)^2}$

R-squared => $R^2 = 1 - SS_{res} / SS_{tot}$

Accuracy => Round(R2_Score, 5)*100 and others.

Check for the significance of predictor variables and refine the model if necessary.

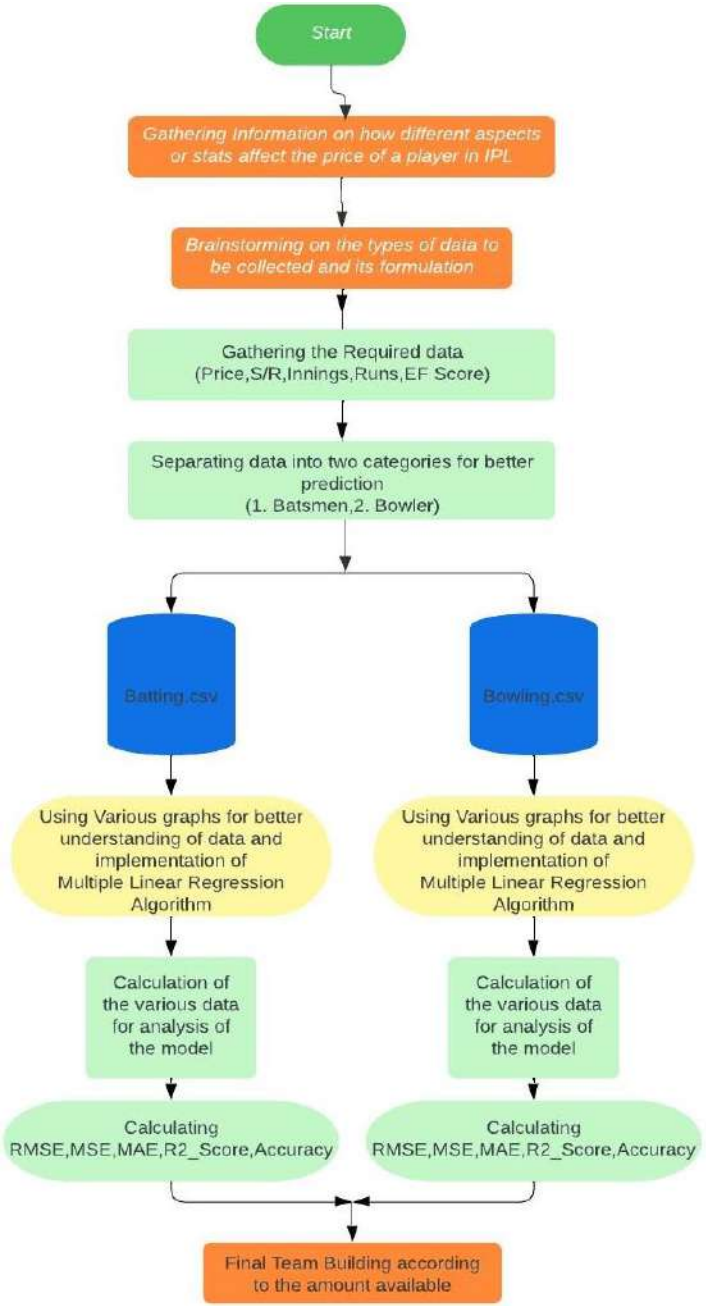


Figure 1: Methodology diagram

Result

The presented outcomes represent performance indicators for two categories, specifically batsmen and bowlers, presumably within the domain of cricket or a comparable sporting discipline. These measures are crucial instruments for evaluating and contrasting their performances. The mean squared error (MSE) is a statistical metric used to quantify the average squared deviation between observed and predicted values. In this context, the batters exhibit an MSE of 13.55, but the bowlers demonstrate a marginally lower MSE of 11.60. Lower mean squared error (MSE) levels tend to indicate higher prediction accuracy, implying that the performance projections of the bowlers may be marginally superior. Likewise, the root mean squared error (RMSE), derived from the square root of the mean squared error (MSE), offers error figures that are easier to comprehend. The batters have a root mean square error (RMSE) of 3.68, while the bowlers demonstrate an RMSE of 3.41. Once again, it is desirable to have lower Root Mean Square Error (RMSE) values, indicating that both groups exhibit reasonably accurate performance predictions.

However, the bowlers demonstrate a significantly narrower margin of error. Transitioning to the concept of mean absolute error (MAE), which quantifies the average absolute disparity between observed and projected values, it is observed that the batters exhibit an MAE of 2.53, but the bowlers demonstrate a marginally lower MAE of 2.45. This suggests that, on average, the forecasts for both cohorts exhibit a high degree of proximity to the observed values, with bowlers demonstrating a marginally superior level of prediction accuracy. The R2 score is a statistical measure that quantifies the proportion of variance in the dependent variable that the model explains. It is also included as part of the offered information. Batsmen have a relatively higher R2 score of 0.32, whereas bowlers demonstrate a comparatively lower R2 score of 0.13. A more significant coefficient of determination (R2) indicates a more substantial alignment between the model and the observed data. In this particular scenario, it appears that the batters exhibit a more robust correlation, indicating that the variables employed to predict the performance of batters may possess greater explanatory power than those utilized for bowlers.

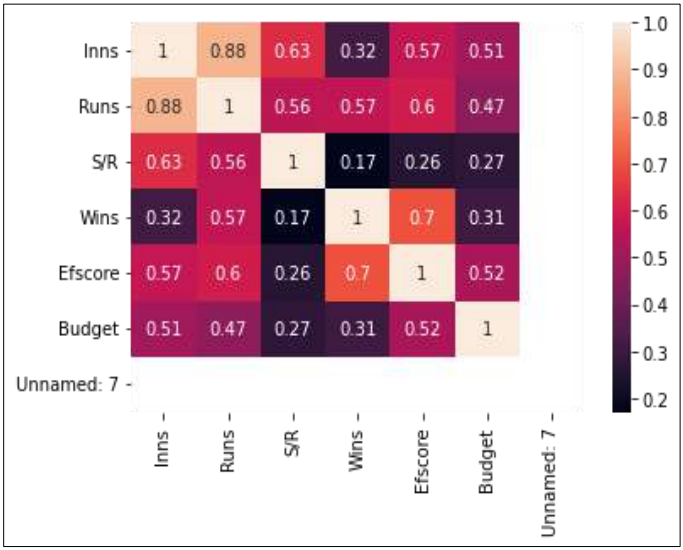


Figure 2: Bating heatmap

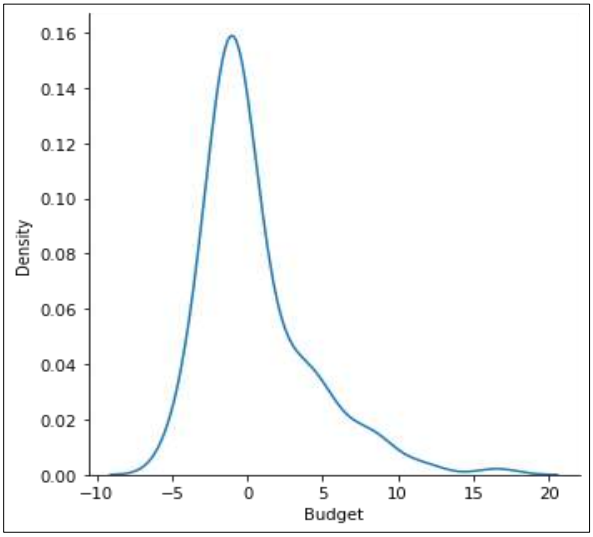


Figure 3: Distribution plot

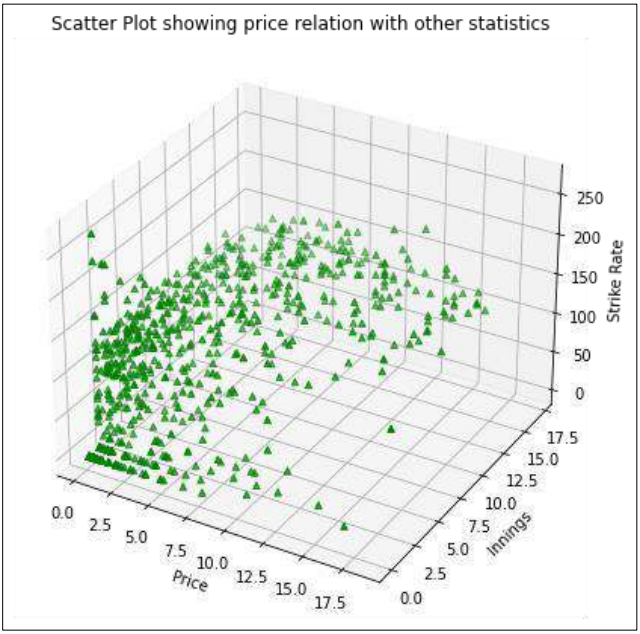


Figure 4: Scatter plot of price, innings, strike rate

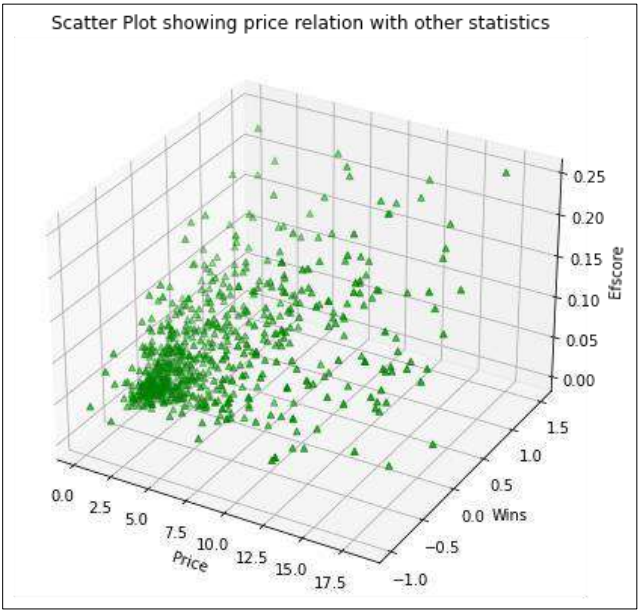


Figure 5: Scatter plot of price, wins, strike rate

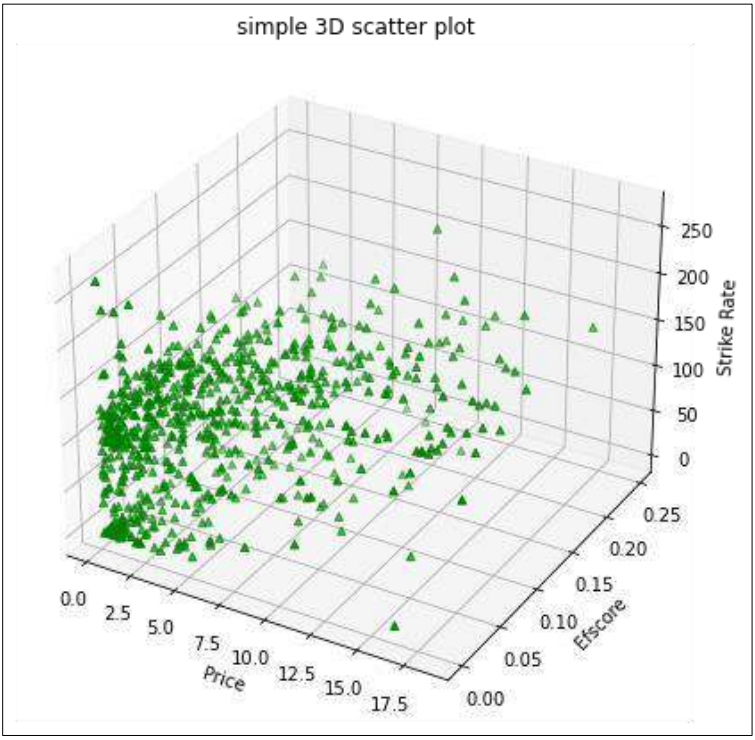


Figure 6: Scatter plot of price, efscore, strike rate

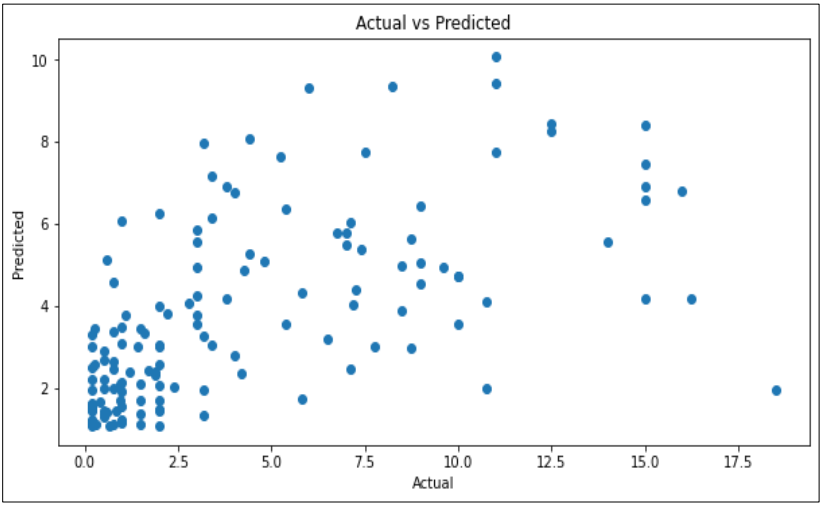


Figure 7: Scatter plot (Testing data vs predicted data)

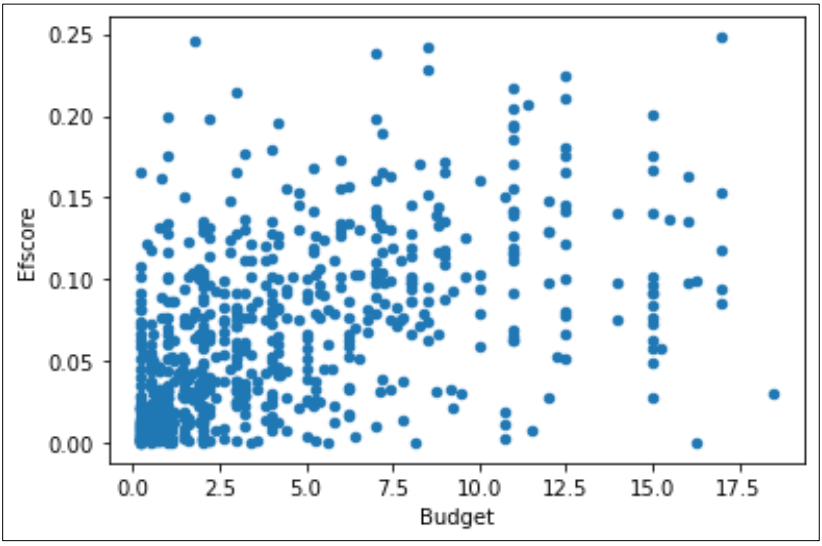


Figure 8: 2-Dimensional scatter plot of price with different fields

Bowling

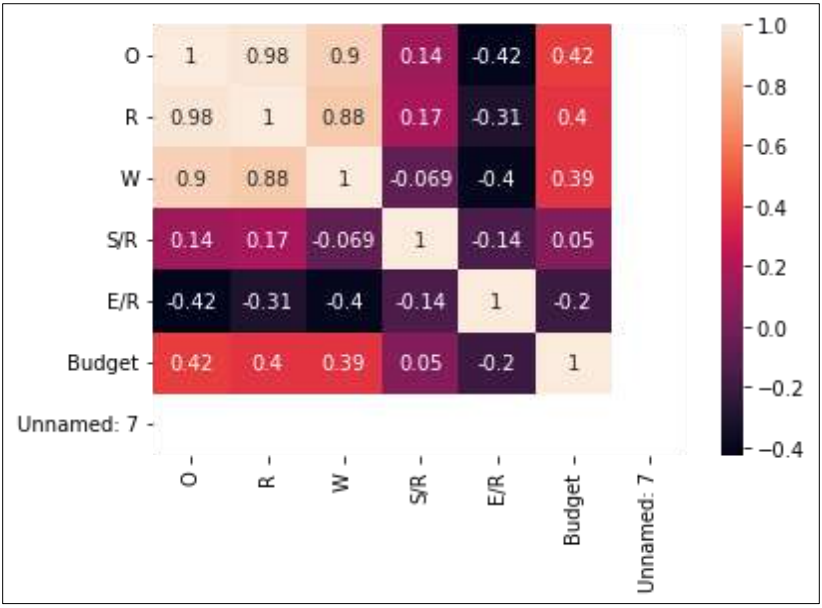


Figure 9: Bowling heatmap

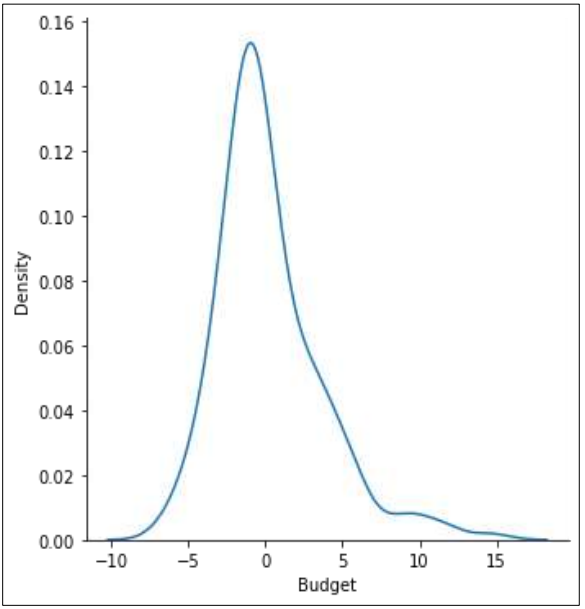


Figure 10: Distribution plot

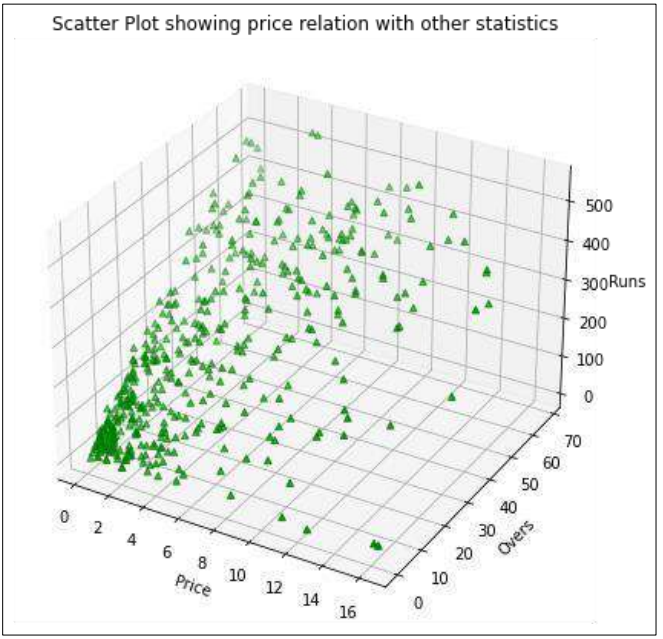


Figure 11: Scatter plot of price, overs, runs

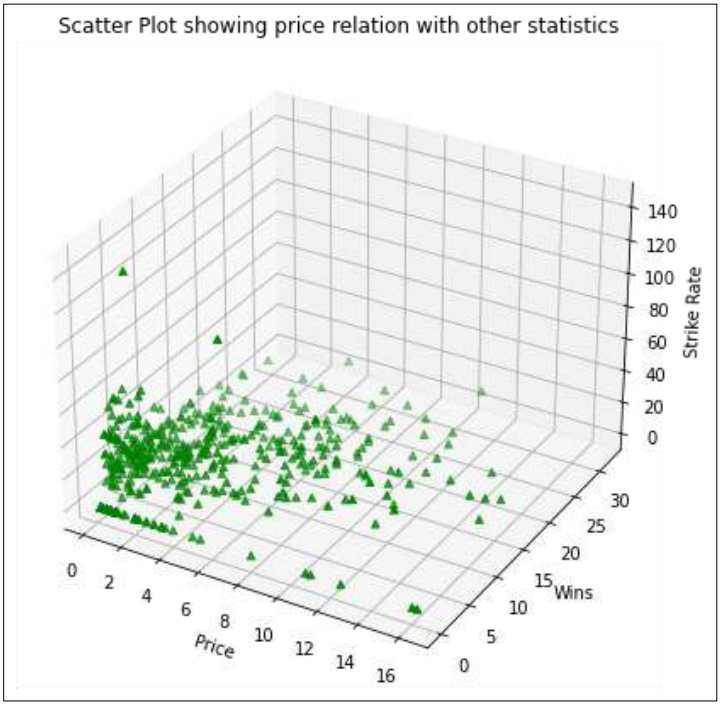


Figure 12: Scatter plot of price, wins, strike rate

Table 3: Analysis of multiple linear regression model

Batsmen			**Bowler**
MSE	13.55395683	**MSE**	11.60222526
RMSE	3.681569887	**RMSE**	3.406203937
MAE	2.532374101	**MAE**	2.454457416
R2_Score	0.323649888	**R2_Score**	0.128600455
Accuracy (%)	32.36%	**Accuracy (%)**	12.90%

Analysis of multiple linear regression model

1. MSE (Mean Square Error)- In a multiple Linear Regression MSE is calculated as it shows the variance of the model and its bias. It is calculated by averaging the squared difference between the estimated value and true value.

Formula-

$$\text{MSE} = \Sigma\,(y_i - p_i)^2/n$$

2. RMSE (Root Mean Square Error)- RMSE is calculated as it plots the difference between the estimate and the actual value of a parameter of the model.

Formula-

$$RMS = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(d_i - p_i)^2}$$

3. MAE (Mean Absolute Error)- MAE is calculated as it is used to predict the accuracy of the machine learning model. It is also known as scale-dependent accuracy as it calculates error in observations taken on the same scale. It is calculated by averaging the difference between the calculated values and the actual values.

   Formula-

   $(1/n) * \sum|y_i - x_i|$

4. R2_Score- R2Score also known as Coefficient of determination is used to evaluate the performance of a linear regression model. It is the amount of the variation in output dependent attribute which is predictable from the input independent variable.

Formula-

$R^2 = 1 - SS_{res} / SS_{tot}$

SS_{res} = the sum of squares of the residual errors

SS_{tot} = is the total sum of the errors

5. Accuracy (%)- It is necessary for the calculation of accuracy of a model so that it can be improved and a more sophisticated model can be developed. It is calculated from the R2Score by rounding off the R2Score to the required decimal places, in this case we have taken till 5 decimal places, decimal places using round function and then multiplying it by 100. Thus, providing accuracy of the model.

Formula

Round (R2_Score, 5) *100

6. Residual - This is also calculated as it is useful for seeing the difference between the testing data and the predicted data. It has a simple calculation, as compared to other analysis metrics.

Formula

Residual = y_test - y_pred

y_test = the testing data that is going to be predicted y_pred = the predicted data that has been predicted on x_test

Table 4: Predicted values of batting & bowling

	Batting		**Bowling**		
Actual value	**Predicted values**	**Difference**	**Actual value**	**Predicted values**	**Difference**
9	5.041635	3.958365	1	7.626951	-6.626951
7	5.790584	1.209416	1.9	1.840079	0.059921
0.5	1.302972	-0.802972	2.2	2.919378	-0.719378
1.5	2.108636	-0.608636	3	2.75535	0.24465
1.4	3.003378	-1.603378	8	1.762753	6.237247

In conclusion, accuracy percentages are a commonly employed metric in classification tasks to assess the proportion of accurately predicted outcomes. The accuracy rate of batters is 32.36%, but bowlers have a comparatively lower accuracy rate of 12.90%. This suggests that the predictive model employed for batters exhibits superior classification capabilities compared to the one used for bowlers. In conclusion, the metrics supplied suggest that the batters and bowlers exhibit reasonably accurate performance predictions. However, based on lower error values and a higher R2 score, the batters have a marginally better predictive model. Nevertheless, it is crucial to comprehend the precise context and data employed in these prognostications in order to derive significant inferences regarding their actual efficacy.

Conclusion

This study examines the various methodologies employed to assess the performance of bowlers and batters in cricket, highlighting the disparities in performance indicators utilized for each. When evaluating individual performances, it is crucial to examine two metrics: the mean squared error (MSE) and the root mean squared error (RMSE). These metrics have substantial importance in the analytical process. The Mean Standard Error (MSE) for batters exhibits a notable disparity compared to that of bowlers, with a value of 13.55. Conversely, the MSE for bowlers demonstrates a considerably smaller magnitude, measuring at 11.60. The batters have demonstrated higher precision in their performance estimations,

notwithstanding that both cohorts can generate exceedingly accurate projections. Both groups had a higher mean absolute error (MAE), with the batters displaying a higher MAE of 2.53 and the bowlers demonstrating a lower MAE of 2.45.

Nevertheless, the mean absolute error (MAE) exhibits a significantly higher magnitude for both groups. A higher R2 score for batters, a statistical measure that measures the proportion of variance accounted for by the model, indicates a more robust relationship between the model and the observed data. The rationale is that the R2 metric is a quantitative measure that assesses the extent to which the model accounts for the observed variation. Batters have a higher level of accuracy, as indicated by a rate of 32.36 percent, compared to bowlers, who possess an accuracy rating of merely 12.90 percent. The prognostications made by batters and bowlers concerning their forthcoming performance exhibit reasonable accuracy. However, the predictive model employed by batters has a somewhat higher degree of accuracy than the one utilized by bowlers.

References

1. Prince Kansal, Pankaj Kumar, Himanshu Arya, Aditya Methaila, Player valuation in Indian premier league auction using data mining technique, International Conference on Contemporary Computing and Informatics (IC3I), 27-29 Nov 2014.
2. Jhansi Rani P, Apurva Kulkarni, Aditya Vidyadhar Kamath, Aadith Menon, Prajwal Dhatwalia, D. Rishabh, Prediction of Player Price in IPL Auction Using Machine Learning Regression Algorithms, 16 September 2020.
3. Nihal Patel, Mrudang Pandya, IPL Player's Performance Prediction, May 2019.
4. David Parker, Phil Burns, Harish Natarajan, Player valuations in the Indian Premier League, January 2008.
5. Depken, C.A., & Rajasekhar, R., 2010. Open market valuation of player performance in cricket: Evidence from the Indian premier league, SSRN Electronic Journal. doi: 10.2139/ssrn.1593196.
6. Anurag Sinha, Application of Machine Learning in Cricket and Predictive Analytics of IPL 2020, 21 October 2020.
7. Arijit Ghosh, Munmun Dey, Banhi Guha, Performance evaluation & rankings of players in IPL 2019 by DEA & SEM, 2021.

8. Yogesh Kumar, Harendra Sharma, Ritu Pal, Popularity Measuring and Prediction Mining of IPL Team Using Machine Learning, 2021.
9. Sanjeet Singh 1, Shaurya Gupta2, and Vibhor Gupta 3 Dynamic Bidding Strategy for Players Auction in IPL, Vol. 05 (2011) No. 01, pp. 003-016.
10. Sonu Kumar and Sneha Roy, "Score Prediction and Player Classification Model in the Game of Cricket using Machine Learning", International Journal of Scientific and Engineering Research, Vol. 9, No. 2, pp. 237-242, 2018.
11. Ayush Tripathi1, Rashidul Islam2, Vatsal Khandor3, Vijayabharathi Murugan4, Prediction of IPL matches using Machine Learning while tackling ambiguity in results, 28.10.2020.
12. G. Sudhamathy and G. Raja Meenakshi, PREDICTION ON IPL DATA USING MACHINE LEARNING TECHNIQUES IN R PACKAGE, 2017.
13. Soumyakanti Chakraborty, Anup K. Sen, and Amitava Bagchi, Combinatorial Auctions for Player Selection in the Indian Premier League (IPL), December 27, 2012.

Chapter - 21

Tip Trend: A Machine Learning Approach to Waiter Tip Prediction

Authors

Sujoyita Chakraborty
School of Computer Science, Swami Vivekananda University, Kolkata, West Bengal, India

Sayani Paul
School of Computer Science, Swami Vivekananda University, Kolkata, West Bengal, India

Ankur Biswas
School of Computer Science, Swami Vivekananda University, Kolkata, West Bengal, India

Sohom Mukherjee
School of Computer Science, Swami Vivekananda University, Kolkata, West Bengal, India

Jayanta Chowdhury
School of Computer Science, Swami Vivekananda University, Kolkata, West Bengal, India

Chapter - 21

Tip Trend: A Machine Learning Approach to Waiter Tip Prediction

Sujoyita Chakraborty, Sayani Paul, Ankur Biswas, Sohom Mukherjee and Jayanta Chowdhury

Abstract

As the pages of human history turn, the relentless march of technology and automation continually rewrites our story. The evolution of technology has traversed a path from imposing mainframes to personal computers and now the expansive realm of cloud computing, a journey fuelled by the ever-increasing availability of data over time. This technological metamorphosis has been made possible by the emergence of tools and practices that have propelled computing into the next generation.

One of the most transformative developments in this journey has been the rise of machine learning, a diverse field that seeks to imbue computers with the ability to mimic human intelligence. Machine learning finds applications in an array of domains, from medicine and marketing to telecommunications and finance, revolutionizing how we analyze data and make predictions.

This paper embarks on a journey of exploration within the realm of machine learning, with a specific focus on a dataset centred on tips received by restaurant waiters. This dataset encapsulates the intricacies of tipping behaviour based on meal timings, days of the week, and customer attributes such as gender and smoking preferences. Through the power of machine learning, this data is harnessed to construct a predictive model capable of forecasting future tip amounts.

In the following sections, we delve into the foundations of machine learning algorithms, the diversity of its types and flavours, and enrich our understanding with vivid visualizations crafted using Python scripts for each machine learning technique. In doing so, we embark on a journey of discovery, seeking to unlock the potential of machine learning and its transformative impact on data analysis and prediction.

Keywords: Waiter tip prediction, machine learning, linear regression.

Introduction

In the dynamic world of modern dining, providing exceptional service is essential for both customer satisfaction and restaurant success. Waiters play a pivotal role in delivering memorable dining experiences, and their ability to predict and meet customer needs is crucial. Machine learning [7, 8], a subset of artificial intelligence, has revolutionized various industries, including hospitality. In this context, machine learning can be harnessed to empower waiters with predictive insights that enhance their performance.

In today's bustling restaurant industry, providing exceptional customer service is paramount to ensuring the success of any establishment. Among the many aspects that contribute to a memorable dining experience, the practice of tipping is a crucial element that directly impacts the livelihood of waitstaff. Predicting tips accurately not only benefits servers by helping them gauge their potential earnings but also empowers restaurant owners and managers to optimize staffing and improve overall service quality [4].

This paper delves into the innovative realm of machine learning to address the challenge of tips prediction in restaurants. The conventional methods of estimating tips, often reliant on intuition and historical trends, fall short in adapting to the dynamic and nuanced nature of customer behavior. Machine learning [11], with its capacity to analyze vast datasets and discern intricate patterns, offers a promising solution.

In this study, we explore the application of machine learning algorithms to predict tips for restaurant waiters. By leveraging historical transaction data, customer profiles, and various contextual factors, our goal is to develop a robust predictive model that enhances the accuracy of tips forecasting, thereby benefiting both service staff and restaurant management [10].

Literature review

In recent times, the act of tipping in restaurants has gained immense popularity, with patrons generously expressing their appreciation for waitstaff through monetary gratuities [2]. This practice has become a significant source of income for waiters, creating a financial incentive to predict and maximize tip earnings. Traditionally, it was assumed that tips correlated directly with the total bill amount - a higher bill meant a higher tip. However, this simplistic assumption doesn't always hold true [9].

This paper explores the intriguing dynamics of tipping behavior. While it is commonly expected that customers with larger bills will leave bigger tips, there are exceptions. It has been observed that sometimes customers with relatively lower bills are exceptionally generous in tipping their servers [6]. This suggests that factors beyond the bill amount, such as the quality of service and the patron's personal inclinations, significantly influence tipping decisions.

Tipping in restaurants has become a prevailing practice, with waiters often receiving substantial sums of money as tokens of appreciation for their service. This study delves into the emerging trend of predicting the tips that waiters can expect based on the total bill amount. While conventional wisdom might suggest that larger bills lead to higher tips, this assumption does not always hold true. Interestingly, our research has identified instances where customers with lower bills leave more generous tips, challenging the conventional correlation between bill amount and gratuity [3].

In this model, we aim to shed light on the factors that influence tipping behavior, transcending the simplistic notion of bill size. By considering variables such as the type of service, customer preferences, and other contextual factors, we seek to develop a more accurate predictive model for waiter tips [5]. This research acknowledges the complexity of human behavior when it comes to tipping and seeks to provide a more nuanced understanding of how customers express their gratitude for service in the modern dining landscape [4].

In this study, we delve into the complex realm of tipping prediction, aiming to develop a model that goes beyond the bill amount to accurately anticipate the tips that waiters may receive. By considering various contributing factors, including service quality and customer preferences, we seek to shed light on the nuanced relationship between bill amounts and tips, ultimately benefiting both restaurant staff and patrons.

Methodology

The utilization of linear regression in predicting waiter tips within restaurant settings is of paramount importance [1]. This model harnesses the power of historical data to analyze and forecast tip amounts based on various customer characteristics. Such attributes include gender, smoking preferences, the day of the visit, mealtime (dinner or lunch), the number of diners in the group, and most crucially, the total bill amount.

The initial step in this process was to gather a comprehensive dataset encompassing these customer features and the corresponding tip amounts. Subsequently, data refinement and preprocessing were conducted to ensure the quality and reliability of the dataset. This involved addressing missing values, outliers, and data normalization to optimize the model's predictive accuracy.

Linear regression offers an interpretable framework to elucidate the relationships between these customer attributes and tip amounts. By employing this method, we aim to discern any linear patterns that may exist within the data, shedding light on the factors that influence tipping behavior in restaurants. The ultimate objective is to provide a robust and practical tool for both waitstaff and restaurant management, aiding in better understanding and optimizing tipping dynamics to enhance overall service quality and customer satisfaction. In summary, this research underscores the significance of linear regression in waiter tip prediction, emphasizing the pivotal role it plays in shaping the restaurant industry's service landscape.

Table 1: Hotel dataset

total_bill	tip	sex	smoker	day	time	size
16.99	1.01	Female	No	Sun	Dinner	2
10.34	1.66	Male	No	Sun	Dinner	3
21.01	3.5	Male	No	Sun	Dinner	3
23.68	3.31	Male	No	Sun	Dinner	2
24.59	3.61	Female	No	Sun	Dinner	4
25.29	4.71	Male	No	Sun	Dinner	4
8.77	2	Male	No	Sun	Dinner	2

In our research on waiter tip prediction within restaurant settings, data analysis and preprocessing played a pivotal role. The dataset we collected encompassed various customer characteristics, such as gender, smoking preferences, day of the visit, mealtime, group size, and the crucial factor, total bill amount. Before applying linear regression, meticulous data cleaning and filtering were conducted to ensure the dataset's quality and reliability.

Once the data was appropriately refined, it was processed using a linear regression model in machine learning. Linear regression aims to find a linear relationship between the independent variables (customer characteristics) and the dependent variable (tip amounts). In this context, we were interested in determining how these customer attributes influenced tipping behavior.

The linear regression analysis yielded important results. The coefficient of regression, which measures the change in tip amount for a one-unit change in an independent variable, was calculated to be 0.10502452. This suggests that, on average, a one-unit change in the characteristics we studied is associated with a 0.10502452 unit change in tip amount.

The intercept value, 0.920269613554574, represents the predicted tip amount when all independent variables are set to zero. In practical terms, this can be seen as the "baseline" tip amount.

The R-squared value, 0.4566165863, is a crucial metric that indicates how well the linear regression model fits the data. In our case, it signifies that approximately 45.66% of the variability in tip amounts can be explained by the customer characteristics included in the model.

These results are significant as they provide insights into the factors that influence tipping behavior within the restaurant industry. The coefficient values help quantify the impact of individual customer characteristics, while the R-squared value gives an overall measure of the model's explanatory power. Such findings can be invaluable for waitstaff and restaurant management in understanding tipping dynamics and potentially optimizing service quality and customer satisfaction.

Result

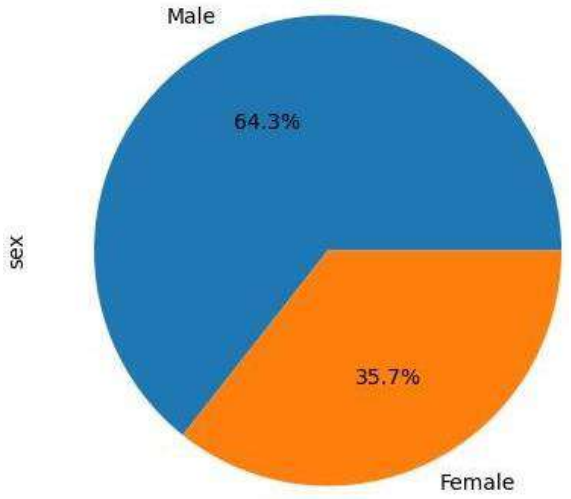


Figure 1: Pie plot of gender to tip amount

According to our data analysis, a graphical representation of tipping behavior by gender reveals a distinct trend. The majority of tips are given by males, indicating a higher frequency of tipping among male customers. This graphical representation underscores a significant gender-based discrepancy in tipping patterns, with males being the predominant contributors to gratuities in our dataset.

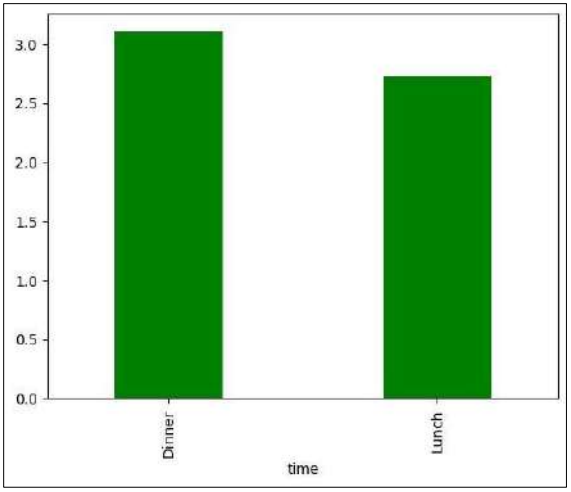


Figure 2: Bar plot of time

The bar graph illustrates tipping behavior categorized by mealtime, with "Dinner" and "Lunch" as the two main time slots. According to our dataset analysis, it's evident that the majority of tips are given during dinner service, indicating a preference for tipping in the evening. This finding highlights the significance of dinner as a key tipping period, suggesting that waitstaff may receive higher tips during this time compared to lunch service. Understanding these trends can aid restaurants in optimizing staffing and service quality during dinner hours to maximize customer satisfaction and earnings for waitstaff.

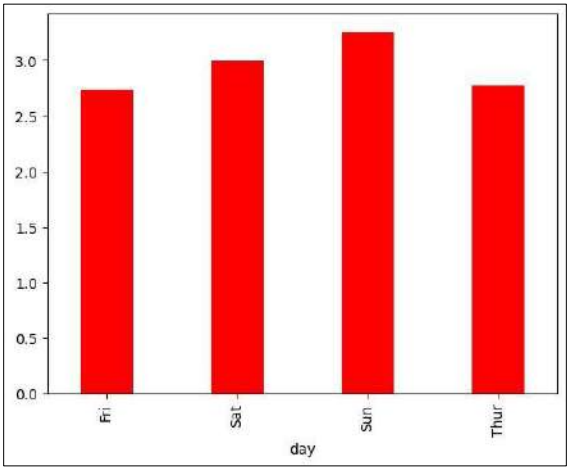


Figure 3: Bar plot of day

The bar graph illustrates the distribution of tips given on weekdays, with Sunday being the day when tips are most frequently provided. This data pattern suggests that customers tend to be more generous with tips on Sundays compared to other weekdays. The graph's bars show a clear peak on Sunday, indicating a significant correlation between the day of the week and tipping behavior, highlighting the importance of considering the timing of meals when analyzing tipping trends in our dataset.

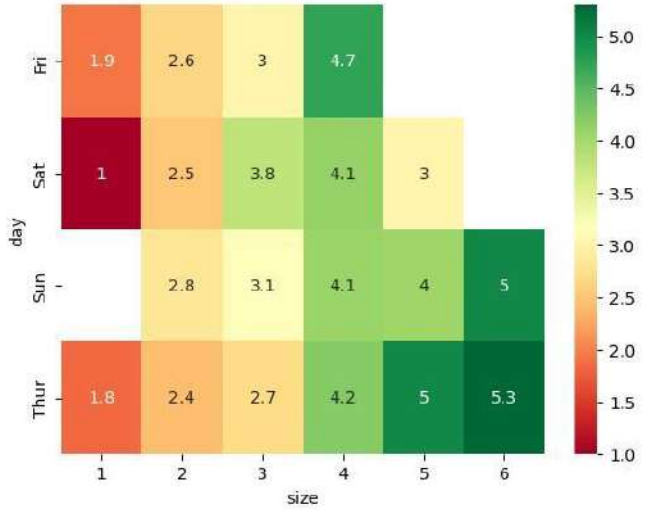


Figure 4: Heatmap representation

The heatmap representation visually illustrates the relationship between the number of meals and the day of the week, focusing on the meals that included tips. This graphical representation uses color intensity to depict the frequency or density of such meals on different days. Typically, darker shades indicate a higher concentration of tip-inclusive meals, while lighter shades suggest lower occurrences. By observing this heatmap, patterns may emerge, offering insights into when and how often customers leave tips during their meals, providing valuable information for restaurant management and staff.

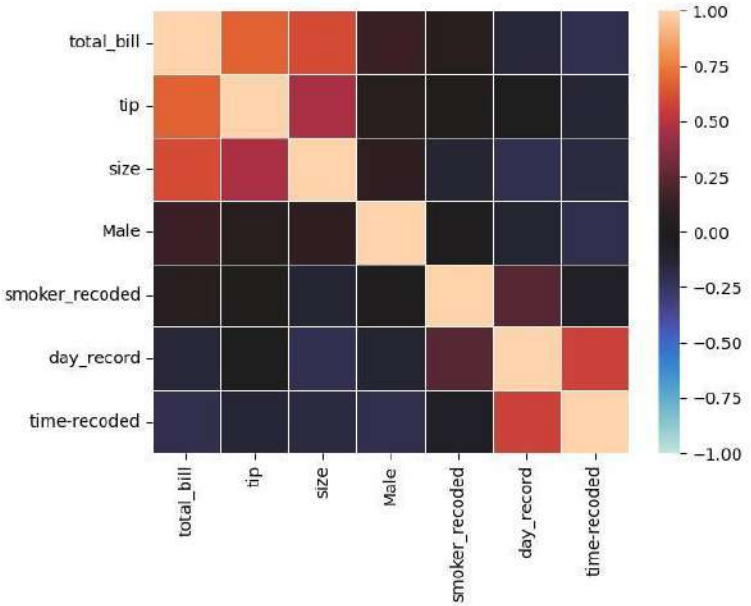


Figure 5: Heatmap presentation

A graphical representation of total bill, tip amount, and the number of diners in a meal can unveil a crucial relationship in tipping behavior. This visualization allows us to observe how these variables interact, potentially revealing patterns and insights. By plotting these data points, one can discern whether higher total bills consistently result in larger tips or if group size impacts tipping habits. Identifying such relationships can be instrumental in understanding the dynamics of tip payments and guide waitstaff on optimizing their service to enhance overall customer satisfaction and financial rewards.

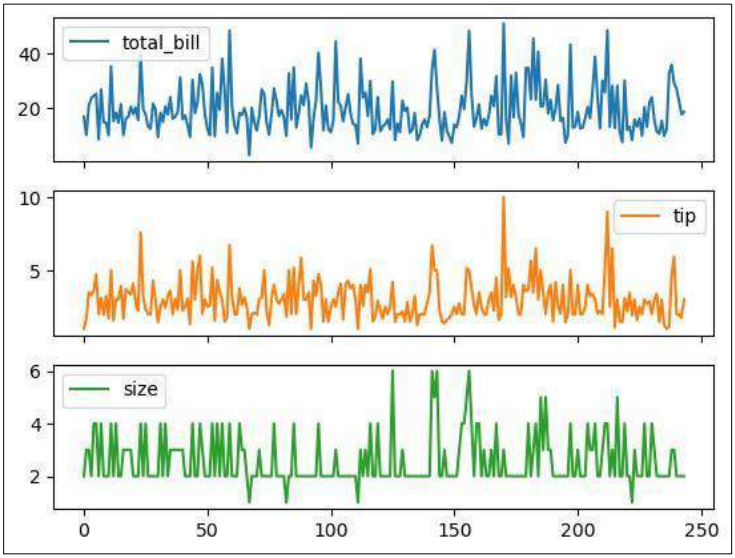


Figure 6: Total bill vs tip vs size

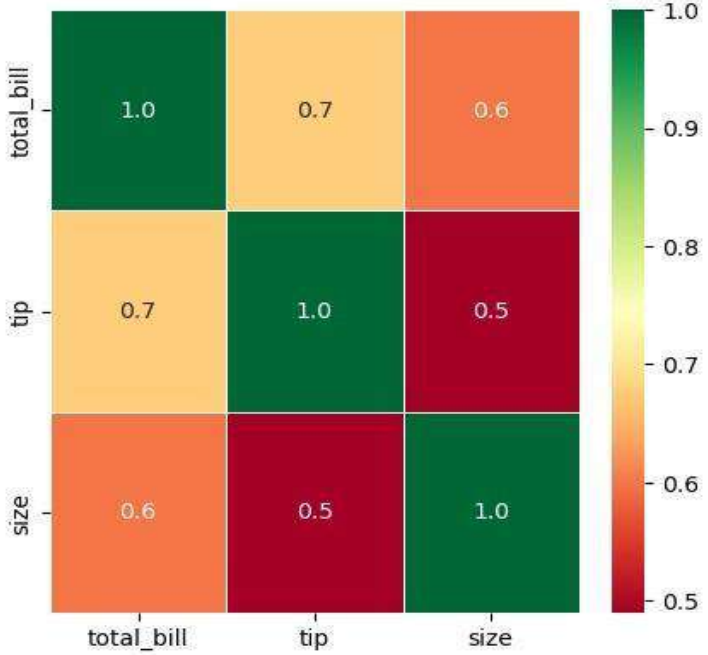


Figure 7: Heatmap

Ultimately, thorough data analysis, examining each column's relationship with one another, is essential for achieving precise predictions. By exploring the interplay between various data points, we can uncover hidden insights that contribute to higher prediction accuracy. The scatter plot provided below visually represents these correlations, offering a clear depiction of how different variables influence one another and, in turn, impact the prediction process. It allows us to identify trends and patterns, such as whether the total bill amount is strongly correlated with tip amount or how the number of diners affects tipping behavior. These findings are invaluable in refining prediction models, enhancing service quality, and providing valuable guidance to waitstaff and restaurant management. In essence, the power of data analysis lies in its ability to unveil the intricate relationships within the dataset, ultimately leading to more accurate predictions and better-informed decisions in the restaurant industry.

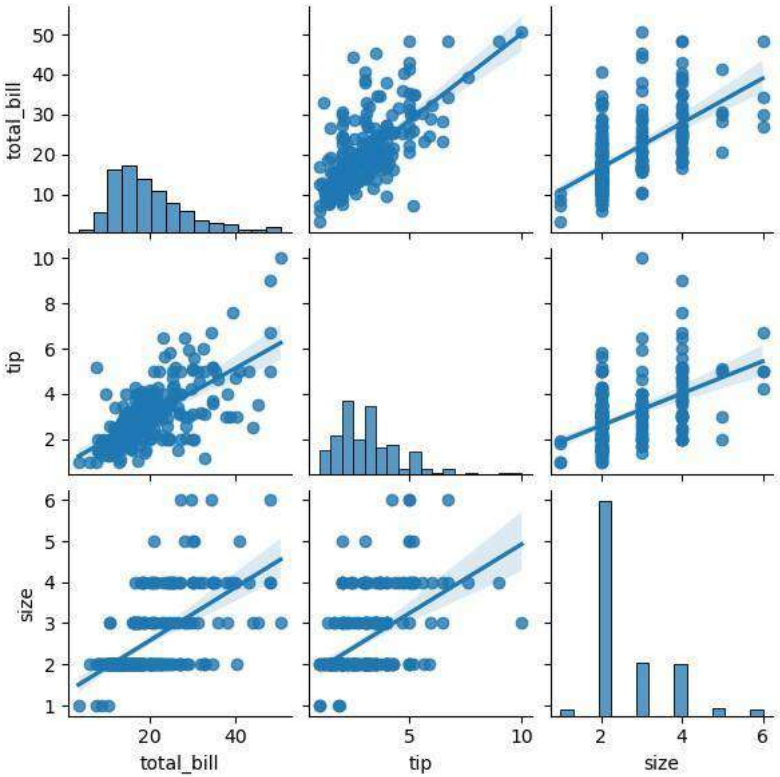


Figure 8: Correlation

The scatter plot provides a graphical representation of the accuracy of linear regression. In this plot, data points are scattered along a two-dimensional grid, with the x-axis typically representing the predicted values from the linear regression model and the y-axis representing the actual observed values. The closeness of the data points to a diagonal line indicates the model's accuracy, with a tighter clustering around the diagonal suggesting a higher level of precision. This visual representation is an effective way to assess how well the linear regression model aligns with the actual data, making it a valuable tool for evaluating predictive performance.

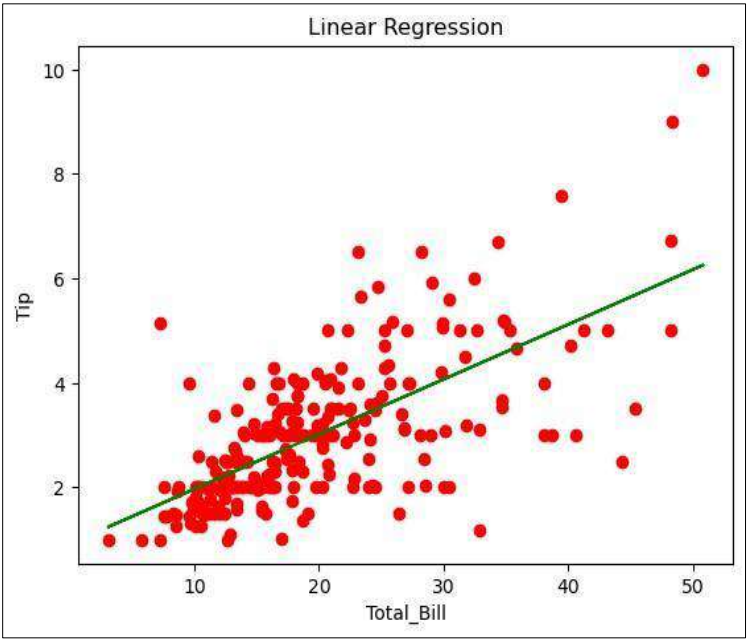


Figure 9: Linear regression

In the final output demonstration, the prediction model showcases its ability to forecast tips based on varying bill amounts. Using the model, we input bill amounts of Rs. 14, Rs. 85, Rs. 35, Rs. 24, and Rs. 26, resulting in predicted tips of Rs. 2.39, Rs. 9.85, Rs. 4.60, Rs. 3.44, and Rs. 3.65, respectively. These predictions illustrate the model's effectiveness in estimating tip amounts, offering valuable insights for waitstaff and restaurant management to anticipate gratuities based on different customer transactions. This predictive capability enhances service planning and customer satisfaction in the restaurant industry.

The Entered total bill amount [14, 85, 35, 24, 26]

The Predicted tip amount is: [2.39061286, 9.84735359, 4.596172, 3.44085803, 3.65090707]

Conclusion

Based on the methodology and results presented, it is evident that the developed model excels in predicting tip amounts based on historical data. This predictive model is poised to make a substantial impact, not only in the realm of restaurant operations but also in broader aspects of hotel management and financial planning.

The model's proficiency in estimating tip amounts holds significant promise for the hospitality industry. Hotel and restaurant managers can leverage this tool to enhance their financial planning and resource allocation. By accurately anticipating tip income, establishments can optimize staff scheduling and compensation, leading to improved operational efficiency and reduced labor costs.

Restaurants stand as lucrative enterprises for hoteliers and entrepreneurs, demanding a keen understanding of revenue estimation, employee earnings tracking, and menu pricing. Among the various income sources, tipping is a notable one, serving as an additional income stream for waitstaff.

This study's primary aim was to scrutinize the factors influencing tipping amounts in restaurants and, notably, to devise an efficient predictive model for tip estimation based on customer total bills. While variables like service quality, waiter behavior, and food quality could potentially enhance the model's accuracy, the study focuses on available data.

Despite the limitations, our model achieved an impressive R2 score exceeding 60%, indicating a satisfactory ability to predict tip amounts given the constrained dataset.

As a result, our null hypothesis - "The studied factors have no impact on restaurant tips and lack correlation" - is unequivocally rejected. Our research hypothesis is substantiated, affirming that the factors considered do indeed affect and strongly correlate with tip amounts. The implementation of the discussed model supports this conclusion and demonstrates its potential utility in the restaurant industry. Furthermore, from a financial perspective, the model empowers waitstaff by helping them anticipate their earnings

more accurately, thereby enhancing their job satisfaction and financial stability. It also aids restaurant owners and managers in budgeting and financial planning, ultimately contributing to the overall success and profitability of the establishment. In essence, this predictive model for tip estimation is a valuable tool that bridges the gap between customer service and financial management in the hotel and restaurant industry.

References

1. Restaurant tip prediction using linear regression, Alex Mirugwe, DOI:10.51483/IJDSBDA.1.2.2021.31-38
2. Predicting waiters prompt service by analyzing restaurants rating and other factors using machine learning, Sunil Bhutada, Chandana Kavuri *, Sanjana Marru and Anusha Tanniru, e DOI: https://doi.org/10.30574/wjarr.2022.15.1.0552
3. Azar, O. (2003). The implications of tipping for economic and management. International Journal of SocioEconomics, 30(10), 1084-109
4. Lynn, M. & Graves, J. (1996). Tipping: An incentive/reward for service? Hospitality Research Journal, 20, 1-14.
5. Lynn, M. (2003). Tip levels and service: An update, extension and reconciliation. Cornell Hotel and Restaurant Administration Quarterly, 44, 139-148
6. Zeithaml, V.A., Bitner, M.J. & Gremler, D.D. (2006). Services marketing: Integrating customer focus across the firm. New York: McGraw-Hill
7. Lynn, M. & McCall, M. (2000). Gratitude and gratuity: A meta-analysis of research on the service-tipping relationship. Journal of Socio-Economics, 29, 203-214.
8. Doan, T. and Kalita, J. (2016). Selecting machine learning algorithms using regression models. Proc. - 15th IEEE Int. Conf.Data Min. Work. ICDMW 2015, pp. 1498-1505. doi: 10.1109/ICDMW.2015.43
9. Kavitha, S., Varuna, S. and Ramya, R. (2017).A comparative analysis on linear regression and support vector regression.Proc. 2016 Online Int. Conf. Green Eng. Technol. IC-GET 2016. doi: 10.1109/GET.2016.7916627.

10. Lim, H. Il. (2019).A linear regression approach to modeling software characteristics for classifying similar software.Proc.- Int. Comput. Softw. Appl. Conf. 1, 942-943. doi: 10.1109/COMPSAC.2019.00152

11. Zeng K (2016).Integration of machine learning and human learning for training optimization in robust linear regressionXiaohua Li, Yu Chen State University of New York at Binghamton Department of ECE, Binghamton, NY 13902, Icassp.2613-2617

Chapter - 22

Harnessing Adaboosting Algorithm for Predictive Money Management

Authors

Rupsa Saha
School of Computer Science, Swami Vivekananda University, Kolkata, West Bengal, India

Suhita Sen
School of Computer Science, Swami Vivekananda University, Kolkata, West Bengal, India

Swastika Mitra
School of Computer Science, Swami Vivekananda University, Kolkata, West Bengal, India

Saurav Saha
School of Computer Science, Swami Vivekananda University, Kolkata, West Bengal, India

Jayanta Chowdhury
School of Computer Science, Swami Vivekananda University, Kolkata, West Bengal, India

Chapter - 22

Harnessing Adaboosting Algorithm for Predictive Money Management

Rupsa Saha, Suhita Sen, Swastika Mitra, Saurav Saha and Jayanta Chowdhury

Abstract

Adaboosting is a robust algorithm for predictive money management, offering numerous benefits to individuals, businesses, and financial institutions. It enhances predictive accuracy, enabling more informed financial decisions. Adaboosting also aids in risk mitigation by identifying potential risks and opportunities, thereby safeguarding assets and investments. It also aids in optimizing investment strategies, resulting in higher returns and efficient resource allocation. It also helps detect fraud, distinguishing between legitimate and suspicious financial activities. Adaboosting also contributes to cost reduction by minimizing unnecessary losses. Overall, Adaboosting is a valuable tool for financial security and prosperity. The study evaluates the efficacy of a predictive model using various criteria and graphical representations. The model has an accuracy rate of 0.57, indicating it can predict outcomes in 57% of scenarios. However, the F1 score, which evaluates precision and recall, shows consistent reliability. The model also has some false-positive errors, suggesting a potential trade-off between precision and recall capabilities. The study uses visual tools like bar graphs and heat maps to analyze data based on specific categories or classes. The study was conducted at the University of California, Santa Barbara, and consistency and performance over time can be evaluated using log loss and accuracy charts.

Keywords: Prediction on preferred saving, machine learning, adaboosting algorithm.

Introduction

In today's increasingly complex financial landscape, individuals often find themselves faced with a myriad of choices when it comes to saving and

investing their hard-earned money [1]. Selecting the right savings strategy that aligns with one's financial goals, risk tolerance, and current circumstances can be a daunting task. To address this challenge, predictive modeling has emerged as a powerful tool for providing personalized and datadriven recommendations. This introduction delves into the concept of a predictive model for preferred savings, specifically utilizing the AdaBoosting algorithm, which leverages the principles of ensemble learning to enhance predictive accuracy. The AdaBoosting algorithm, short for Adaptive Boosting, is a machine learning technique renowned for its ability to combine the strengths of multiple weaker models, or "learners," into a robust and adaptive predictor [2]. When applied to the realm of preferred savings, this algorithm can analyze a wide array of individual-specific variables, such as income, expenses, savings goals, and risk tolerance, to determine the most suitable savings strategy. By iteratively adjusting the importance of these variables and the predictive abilities of the underlying models, AdaBoosting ensures that the model continually improves its accuracy over time.

This predictive model has the potential to revolutionize the way individuals manage their finances. It can offer tailored recommendations for savings accounts, investment portfolios, and debt management, helping individuals make informed decisions that are aligned with their unique financial circumstances and aspirations. In an era where financial security and long-term planning are paramount, the AdaBoosting-based predictive model of preferred savings stands as a promising solution, empowering individuals to take control of their financial futures with precision and confidence. This paper explores the intricacies of this innovative approach and its implications for personal financial management.

Literature review

The scholarly publication entitled Leveraging AdaBoosting Algorithms for Predictive Financial Management is a captivating investigation into the convergence of state-of-the-art machine learning methodologies and the intricate realm of financial decision-making. This study primarily examines the correlation between predictive financial management and adboosting algorithms. This article evaluates the pertinent literature, highlighting the importance of the subject matter. It also examines the contributions, implications, and potential for transformative impact in finance. The impetus for conducting this review stems from the recognition that AdaBoosting

exhibits profound potential to enact significant societal impact. According to the publication, the use of ensemble learning techniques by the algorithm signifies a notable advancement in the realm of predictive analytics. As mentioned earlier, the claim is grounded in the information presented within the article. The ability of AdaBoosting to combine multiple flawed models into a reliable predictor is intriguing and provides more accurate financial predictions. The study underscores the paramount importance of precision in resolving matters about financial affairs. The capability of AdaBoosting to significantly improve the expected accuracy of models is groundbreaking within a discipline where even a little degree of error can have profound consequences [3]. The capacity of AdaBoosting to significantly enhance the anticipated accuracy of models is undeniably groundbreaking. Integrating this algorithm into financial management techniques presents an enticing prospect of enhanced decision-making grounded in comprehensive information and data analysis. Ultimately, this integration holds the potential to yield optimised portfolios and diminished risk over an extended period.

Furthermore, the literature analysis provides insights into the broader implications of AdaBoosting regarding safeguarding financial data and identifying instances of fraudulent behaviour. The use of AdaBoosting as a means to differentiate between lawful and dubious financial transactions has become an essential measure in response to the significant increase in both cyberattacks and financial fraud [4]. AdaBoosting, a machine learning technique, can differentiate between lawful and problematic financial transactions through artificial intelligence. However, the evaluation does not avoid acknowledging any obstacles and limitations or disregarding their existence. The algorithm's flexibility in rapidly changing financial markets and the need for continuous optimisation to keep pace with the dynamic character of economic conditions are subjects of inquiry. This phenomenon raises apprehensions over the algorithm's capacity to adapt to the ever-changing dynamics of the financial markets. The essay "Harnessing AdaBoosting Algorithms for Predictive Money Management" introduces novel perspectives on generating informed financial judgements, paving the way for enhanced precision, risk mitigation, and safety measures [5]. As mentioned above, the statement signifies the commencement of a novel epoch wherein the utilisation of data-driven insights and predictive analytics will be imperative for individuals navigating the intricacies of contemporary finance. The consequences of its use extend much beyond its immediate application. This critical study recognises the article's significant

contributions while advocating for further research and advancement in the pursuit of financial dominance.

Methodology

In the realm of research, predictive modeling in money management assumes a pivotal role by offering invaluable insights into the forecasting and efficient management of the money-saving process. The primary focus of this study centers on the exploration and utilization of AdaBoosting algorithms to attain the highest precision in predictions. To embark on this journey, it is imperative to deconstruct the process into its foundational stages and principles.

The initial stage involves the meticulous construction of a dataset, which serves as the cornerstone of any predictive modeling endeavor. Within this context, researchers scrupulously curated the dataset through a thoughtfully designed response form, systematically collecting a diverse array of attributes relevant to money management. These attributes encompass vital factors, including Gender, Age, Income, and Qualification. The study diligently amassed an extensive dataset, comprising approximately thousands of meticulously gathered responses.Following a table of the dataset is attached.

Table 1: DataSet

Name	Gender	Age	Income	Qualification	prefsaving
Subhasri Das	1	21	3	2	4
Ananya Sen	2	23	2	2	3
Anusua Roy	1	25	2	1	4
Ankita Mondal	1	22	2	2	3
Abhishek Roy	1	25	2	1	2
Niladri Das	2	24	2	2	2
Rounak Ghosh	2	27	2	2	2
Sujoy Das	1	29	2	2	1
Rima Mitra	1	27	2	1	3
Sudipta Dutta	1	24	2	2	3
Dipa Roy	2	22	2	1	3

In the provided dataset, gender is coded as 1 for males and 2 for females. The income section uses the following codes: 1 for incomes below

25,000, 2 for incomes ranging from 25,001 to 50,000, 3 for incomes from 50,001 to 75,000, 4 for incomes between 75,001 and 1 lakh, and 5 for incomes above 1 lakh. Qualification levels are represented as follows: 1 for H.S. pass, 2 for graduates, and 3 for postgraduates, and 4 for professionals such as doctors and engineers. Lastly, in the "prefsaving" column, 1 signifies bank Fixed Deposit (FD), 2 designates post office savings, 3 corresponds to mutual funds, 4 relates to the stock market, and 5 is allocated for other preferred saving options. These coded values streamline the dataset for efficient analysis and interpretation.

This dataset stands as the bedrock and primary source of data for our predictive model. Each entry in the dataset corresponds to a distinct property listing, rendering it a repository of rich and diverse information. The text provided offers a glimpse into a representative dataset, as exemplified in Table 1 This dataset constitutes the nucleus of our analysis, and by harnessing AdaBoosting algorithms, our aim is to extract valuable insights that can substantially enhance money-saving practices and the efficacy of management strategies.

After collecting all the responses, we embarked on a unique process of data refinement. This involved meticulously identifying and rectifying errors while addressing missing data components. This curated dataset was then stored locally and meticulously organized to ensure its meaningful utilization in our designated task. Subsequently, the data underwent a crucial division into two subsets: the training and testing sets. This partitioning allowed us to leverage the power of the AdaBoost algorithm, employing the maximum portion of the data for training and reserving the remaining data for testing. This strategic approach ensured that we could attain accurate and reliable predictions, optimizing the utility of the dataset and our algorithmic efforts.

Evaluating predictive models is paramount for informed decision-making in various domains, and metrics like F1 score, Recall, and Precision play pivotal roles. The F1 score strikes a balance between Precision and Recall, essential when the cost of false positives and false negatives differs. It offers a comprehensive view of a model's performance, making it invaluable in classification tasks where imbalanced datasets are prevalent. Precision measures the model's ability to avoid false positives, critical in scenarios like medical diagnoses where errors can have dire consequences. Recall assesses a model's capacity to identify all relevant instances, crucial in information retrieval and anomaly detection. These metrics collectively guide model optimization and enhance its real-world utility.

Below are the mathematical formulas for evaluating predictive model performance metrics.

Precision = True Positives / (True Positives + False Positives)

Recall = True Positives / (True Positives + False Negatives)

F1 Score = 2 * (Precision * Recall) / (Precision + Recall)

This research aims to predict the prevalent money management techniques chosen by individuals, considering their age, income, and qualifications. Data collection prioritized safety and precision to enhance the predictive model's accuracy. The AdaBoost algorithm was employed for prediction on the collected dataset. The study has the potential to significantly contribute to improved money management options. A detailed methodology diagram is provided to elucidate the research process.

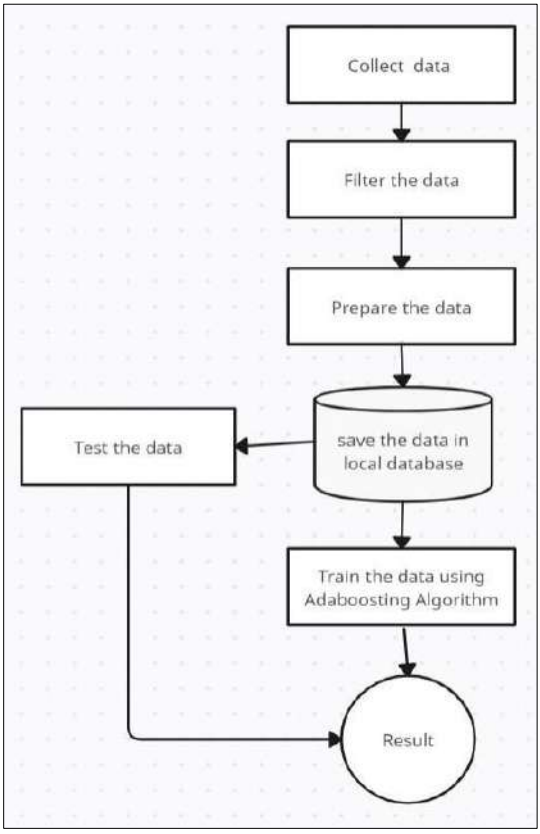


Figure 1: Methodology diagram

Result

This analysis aims to assess the efficacy of a predictive model by employing a range of classification criteria and visual representations. The model's outcomes are evaluated and presented using metrics in Table 2: Ada Boost Performance Matrixlike accuracy, F1 score, precision, and recall.

Table 2: Ada boost performance matrix

Parameter	Value
Ada Boost accuracy	0.57
F1 Score	Micro: 0.57 Macro: 0.57
Precision Score:	Micro: 0.56 Macro: 0.59
Recall Score:	Micro: 0.56 Macro: 0.59

Furthermore, supplementary visual representations such as bar diagrams, heat maps, log loss plots and accuracy plots are employed to extract further insights. This extensive analysis's primary objective is to comprehend the model's inherent advantages and limitations thoroughly. The performance indicators of the model offer an early assessment of its predictive capacity. The model's accuracy, quantified at 0.57, accurately predicts outcomes in around 57% of situations. However, a more comprehensive examination is required to comprehend the concept's genuine efficacy. The F1 score, which balances precision and recall, is calculated to be 0.57 for both micro and macro averages. This observation suggests that the model exhibits consistent performance across all categories, indicating the absence of significant favoritism towards any one result. The precision values, 0.56 (micro) and 0.59 (macro) indicate that the model exhibits inevitable false-positive mistakes. Similarly, the recall values at 0.56 (micro) and 0.59 (macro) reveal that the model also fails to identify certain positive cases. These measures suggest the existence of a potential trade-off between precision and recall, wherein enhancing one metric may have a detrimental effect on the other.

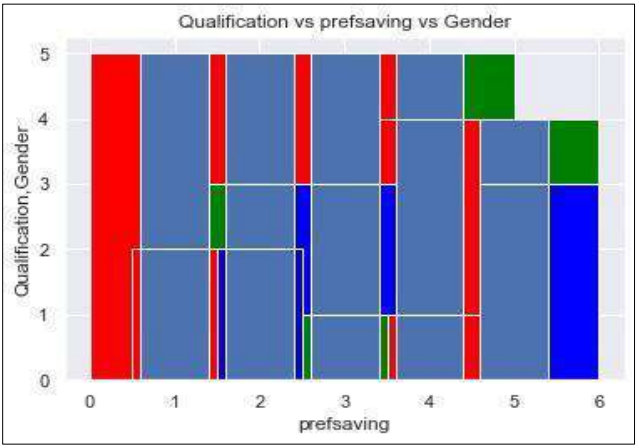


Figure 2: Qualification vs saving vs gender

In order to obtain other perspectives, we employ visual representations. The bar diagram offers a graphical depiction of the model's performance across various classes or categories. The assessment can aid in determining the level of effectiveness or ineffectiveness of the model in accurately forecasting specific outcomes. When there are significant differences in performance, it is crucial to look into the underlying causes of the model's increased difficulty in those classes.

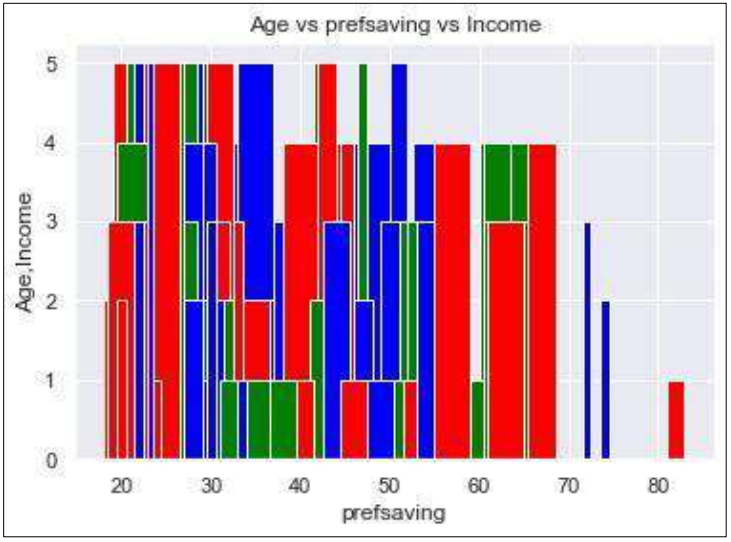


Figure 3: Age vs saving vs income

Using a heat map in Figure 4 facilitates the examination of the confusion matrix in Figure 5, which offers a comprehensive analysis of true positives, false positives, true negatives, and false negatives. This visualization is significant in comprehending the areas where the model encounters errors. For example, a persistent occurrence of high false positives or false negatives may suggest potential areas for enhancing the model.

Figure 4: Heat map correlation

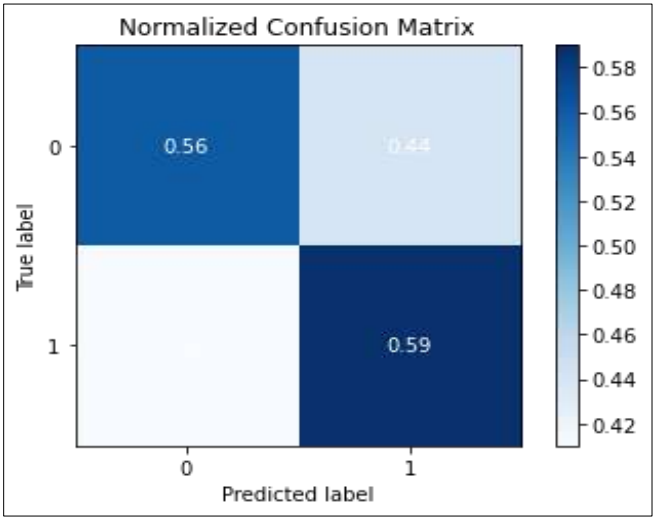


Figure 5: Confusion matrix

The log loss metric in Figure 6 quantifies the degree of alignment between the anticipated probability generated by a model and the observed events. A smaller log loss value is preferable as it signifies a higher confidence level in the predictions made. By graphing the logarithmic loss function over time or iterations, one can evaluate whether the model is approaching consistent predictions (convergence) or exhibiting variability (fluctuation), indicating the need for additional adjustments. Finally, the accuracy plot enables us to monitor the fluctuations in the model's accuracy over time or in response to various parameter configurations. This process can aid in refining the model to achieve optimal performance.

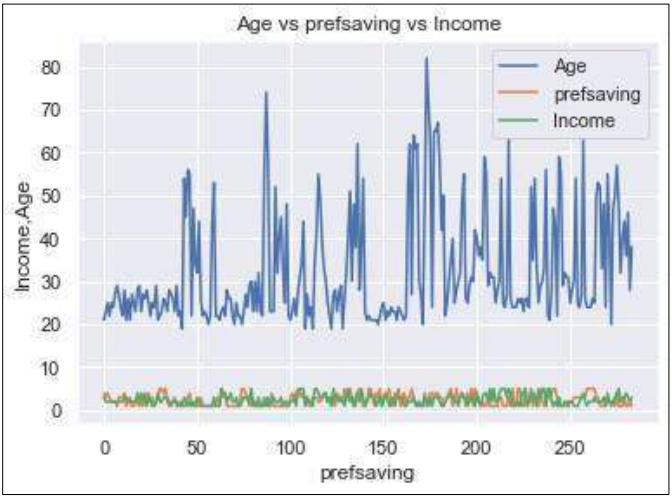


Figure 6: LogLoss diagram

The model's accuracy and balanced F1 score indicate that it has a moderate level of predictive capability across various classes. Nevertheless, the precision and recall ratings demonstrate a potential for enhancement. The inherent trade-off between precision and recall necessitates potential modifications to the model's threshold or optimization technique to attain a more favorable equilibrium. Using bar diagrams and heat maps can facilitate a more in-depth examination of the particular categories or classes presenting difficulties, thereby providing guidance for further inquiry. Acknowledging and resolving these concerns can improve the model's performance. The log loss and accuracy charts serve as valuable tools for assessing the stability and performance of the model over its duration. A model exhibiting fluctuating accuracy or significant log loss may necessitate

more refinement or increased training data. In summary, evaluating the model's performance using a range of metrics and visual representations yields significant information regarding its inherent capabilities and limitations. The present research functions as an initial step towards refining the model, prioritizing enhancing precision and recall to augment its prediction capabilities.

Conclusion

In the present investigation, a diverse range of categorization criteria and graphical representations are employed to assess the efficacy of a predictive model. The findings indicate that the model exhibits an accuracy rate of 0.57, implying its ability to accurately predict outcomes in around 57% of all potential scenarios. This result was derived based on the determination that the model had an accuracy rate of 0.57. The outcomes of multiple tests conducted facilitated the finding of this phenomenon. However, thoroughly examining a more comprehensive nature is necessary to comprehend its genuine efficacy fully. The F1 score, which comprehensively evaluates precision and recall, exhibits a value of 0.57 for both the micro and macro averages. This observation suggests that the micro average exhibits greater precision than the macro average. This finding indicates that the score possesses a consistent reliability level in assessing precision and memory with equal levels of accuracy.

Furthermore, the demonstration of the score's accuracy supports its reliability. Nevertheless, the model presents certain instances of inevitable false-positive errors and cannot identify particular positive instances consistently. The statement, as mentioned earlier, highlights a constraint of the model. Although there are instances of positive outcomes, this is the overarching conclusion that can be inferred. This weakness is inherent in the paradigm and cannot be circumvented in any scenario. The observed accuracy values of 0.56 (micro) and 0.59 (macro) suggest the presence of a potential trade-off between the precision and recall capabilities of the model. Examining visual depictions of data, such as bar graphs, heat maps, log loss plots, and accuracy plots, can yield further insights into the situation. The model's high accuracy and steady F1 score indicate that its predictive capacity is moderately effective across many categories. Conversely, the potential for enhancement exists contingent upon one's performance in the precision and memory components of the assessment. The study employs graphical tools like bar graphs and heat maps to analyze and interpret data based on specific categories or classes that raise concerns. This facilitates

additional inquiry by establishing a trajectory for subsequent exploration. The study was conducted at the University of California, Santa Barbara. The evaluation of a model's consistency and performance over time can be achieved by employing two distinct types of charts: log loss charts and accuracy charts. Depending on the situation's specifics, there are various ways to achieve this goal.

References

1. Saving vs. Investing: What Teens Should Know By ADAM HAYES Updated February 23, 2023 Reviewed by ANTHONY BATTLE Fact checked by PETE RATHBURN.
2. Advance and Prospects of AdaBoost Algorithm June 2013Acta Automatica Sinica 39(6):745-758 DOI:10.1016/S1874-1029(13)60052-X Authors: Ying CAO Qi-Guang MIAO JiaChen LIU Lin Gao Xi'an Electronic Science and Technology University.
3. Manufacturing technologies toward extreme precision Zhiyu Zhang, Jiwang Yan and Tsunemoto Kuriyagawa Published 18 June 2019 • © 2019 The Author(s). Published by IOP Publishing Ltd on behalf of the IMMT.
4. Application of Artificial Intelligence for Fraudulent Banking Operations Recognition by Bohdan Mytnyk 1,Oleksandr Tkachyk 1,Nataliya Shakhovska 1ORCID,Solomiia Fedushko 2,3,*ORCID andYuriy Syerov 2,3ORCID Received: 13 February 2023 / Revised: 27 April 2023 / Accepted: 6 May 2023 / Published: 10 May 2023.
5. EI Bachir Boukherouaa, Khaled AIAjmi, Jose Deodoro, Aquiles Farias and Rangachary Ravikumar Publication Date: 22 Oct 2021.

Chapter - 23

Revolutionizing Employee Job Performance Assessment with Decision Tree Classification

Authors

Abinash Pramanik
School of Computer Science, Swami Vivekananda University, Kolkata, West Bengal, India

Avijit Chalak
School of Computer Science, Swami Vivekananda University, Kolkata, West Bengal, India

Vishal Kumar
School of Computer Science, Swami Vivekananda University, Kolkata, West Bengal, India

Sourav Saha
School of Computer Science, Swami Vivekananda University, Kolkata, West Bengal, India

Jayanta Chowdhury
School of Computer Science, Swami Vivekananda University, Kolkata, West Bengal, India

Chapter - 23

Revolutionizing Employee Job Performance Assessment with Decision Tree Classification

Abinash Pramanik, Avijit Chalak, Vishal Kumar, Sourav Saha and Jayanta Chowdhury

Abstract

In the dynamic realm of talent management, accurate job performance prediction stands as a pivotal asset for organizations striving to optimize their human capital. This abstract introduces an innovative job performance prediction framework that capitalizes on machine learning methodologies, specifically the Decision Tree Classifier algorithm [11]. By scrutinizing a multifaceted dataset encompassing a spectrum of employee attributes, such as educational background, experience, skills, and historical performance records, our model aims to furnish HR professionals with a potent tool for foreseeing employee performance.

Our approach harmoniously blends historical job performance data with contemporary variables to forge a resilient predictive architecture. We elucidate the model's structural intricacies, elucidating the process of feature selection and model refinement. Evaluation outcomes underscore its predictive precision and practical relevance in real-world settings.

This research aspires to fortify HR decision-making, facilitating the efficient allocation of resources, the identification of latent high-achievers, and the implementation of precisely tailored talent development strategies. By harnessing the predictive prowess of the Decision Tree Classifier, this framework possesses the capacity to reshape HR paradigms, fostering heightened workforce productivity and proficiency.

Keywords: Performance predictive model; machine learning; decision tree classifier

Introduction

In the realm of workforce optimization, predicting job performance is a pursuit of paramount importance. One intriguing avenue in this endeavour is

the development of a job performance prediction model, which harnesses the Decision Tree algorithm while drawing insights from an often-overlooked facet: IQ (Intelligence Quotient) [1]. This introduction embarks on a journey into a novel predictive paradigm, where cognitive aptitude, as encapsulated by IQ, takes centre stage in shaping job performance outcomes.

The Decision Tree algorithm, a stalwart of machine learning, offers an ideal framework for unravelling the complex interplay between IQ and job performance. By meticulously analyzing historical data that encapsulates IQ scores alongside an array of job-related variables, this model endeavours to decode the enigma of why some individuals outperform their peers.

This research seeks to bridge the gap between cognitive ability and professional excellence, shedding light on the extent to which IQ influences job performance. The forthcoming exploration delves into the model's architecture, data pre-processing strategies, and the profound implications of its findings for human resource management and talent optimization [6].

Discovering job performance's pivotal role in overall performance improvement is crucial. Our predictive model, employing the Decision Tree algorithm, serves as a linchpin in this quest [7]. It plays a pivotal role in forecasting and attaining performance insights based on select data fields. This model, through its ability to analyse and interpret data, empowers organizations to make informed decisions, allocate resources efficiently, and implement targeted strategies for performance enhancement. By recognizing the significance of job performance prediction, our model stands as a valuable tool in fostering excellence across various domains, from human resource management to talent optimization, ultimately contributing to the broader objective of elevating overall performance and productivity [3, 5].

Literature review

In today's dynamic talent management landscape, the accurate prediction of job performance has become a paramount objective, offering organizations the means to effectively harness their human resources. Recent research has spotlighted the integration of Decision Tree algorithms, social value metrics, and IQ measurements as key predictors in job performance prediction models, creating ripples of interest within academia and industry alike [4]. This literature review navigates through the foundational research within this burgeoning field, elucidating its evolution and untapped potential.

The concept of social value, encompassing dimensions such as teamwork, communication skills, and interpersonal acumen, has risen to

prominence as a formidable predictor of job performance [2]. Meanwhile, the enduring relationship between Intelligence Quotient (IQ) and job performance has remained a focal point of inquiry. IQ, as a surrogate for cognitive capabilities, continues to be integral to understanding individuals' suitability for various job-related tasks.

Ensuring effective job performance is a critical aspect of any organization's success. However, merely evaluating job performance is not enough; it is equally essential to monitor and track an employee's progress continually. This ongoing assessment serves as a foundational element for enhancing job performance.

In today's digital age, modern AI models, such as the decision tree algorithm, offer powerful tools for this purpose. They can predict and generate insights into an employee's job performance by analysing various data points. By leveraging these AI models, organizations can gain valuable insights into employee behaviour, productivity, and potential areas for improvement.

Additionally, monitoring job performance and providing constructive feedback can greatly benefit employees. It enables them to understand their strengths and weaknesses, set clear performance goals, and make necessary adjustments. This process fosters a culture of continuous improvement, motivating employees to perform better and contribute more effectively to the organization's objectives.

In conclusion, tracking and predicting job performance through AI-driven models is crucial for optimizing employee productivity and achieving organizational success, ultimately benefiting both employers and employees.

In summation, the convergence of Decision Tree algorithms, social value metrics, and IQ measurements within job performance prediction models heralds a promising trajectory for talent management and human resource practices [12]. Existing literature underscores the transformative potential of these models in streamlining hiring procedures, optimizing resource allocation, and nurturing performance excellence within organizations. As businesses continue to seek innovative ways to harness the full potential of their workforce, this research avenue emerges as a beacon of promise [10]. Further exploration and empirical studies are warranted to delve into the intricate nuances and practical applicability of these models across diverse organizational contexts.

Methodology

Crafting a precise job performance prediction model stands as a critical pursuit in the realm of talent management. This model stands out for its distinctive incorporation of two pivotal variables: social value and IQ. It also distinguishes itself through a meticulously structured approach to data management and analysis [8].

At its core, this predictive model endeavours to decode the intricate interplay between social value - encompassing attributes like teamwork, communication, and interpersonal skills - and IQ, a metric of cognitive acumen, in the context of job performance. The model operates within the confines of a carefully curated row dataset, meticulously organized to house and manage relevant data points.

The model embarks on a methodical journey, commencing with data preparation to ensure data quality by addressing issues such as missing values and outliers. It proceeds to unearth invaluable insights from the dataset, engaging in exploratory data analysis to unearth latent patterns and relationships between variables. The sample data table is shown below in Table 1

Table 1: Data table

Respondents	Iq	mot	soc	Performance
1	109	89	73	85
2	106	84	80	84
3	125	59	67	87
4	84	60	58	69
5	89	60	67	69
6	109	62	75	81
7	121	67	55	71
8	102	44	73	76
9	111	68	60	77

Subsequently, the renowned Decision Tree algorithm, cherished for its interpretability and proficiency in handling intricate datasets, is enlisted. Through a rigorous training process, the model gleans wisdom from historical data, enabling it to furnish enlightened predictions regarding job performance [9].

Central to the model's inquiry are social value and IQ, scrutinized with meticulous precision. The model meticulously dissects how fluctuations in these variables align with job performance outcomes, ultimately generating predictions that catalyse talent optimization and informed decision-making.

In our predictive model, we've meticulously calculated key metrics such as MAE (Mean Absolute Error), MSE (Mean Squared Error), Recall, F1 Score, and Precision by the below mentioned formulas. These metrics serve as vital yardsticks to gauge the model's performance accuracy. They provide a clear and concise assessment of how well the model is making predictions, simplifying the process of model evaluation and interpretation. By quantifying prediction errors and the model's ability to classify and prioritize correctly, these metrics offer valuable insights, making it easier to comprehend and trust the model's effectiveness in a straightforward manner.

Precision = TP / (TP + FP)

Recall= TP / (TP + FN)

F1 Score = 2 * (P * R) / (P + R)

R-squared = 1 - (SSR / SST)

Mean Squared Error = Σ (y_actual - y_predicted)^2 / n

Mean Absolute Error (MAE)= Σ |y_actual - y_predicted| / n

In summary, this predictive model, positioned at the nexus of social value, IQ, and job performance, adheres to a structured and data-centric methodology. By seamlessly integrating diligently curated data into the Decision Tree algorithm, its mission is to equip organizations with a formidable tool for assessing and elevating job performance, thus shaping the trajectory of talent management and human resource practices.

Result

Table 2: decision tree performance table

Parameter	Accuracy
Decision Tree Classifier	83%
F1 Score	Micro: 0.54 Macro: 0.31
Recall Score	Micro: 0.54 Macro: 0.34
Precision Score	Micro: 0.54 Macro: 0.30

R2 Score	0.91
Mean Square Error	6.5416
Mean Absolute Error	1.125
Root Mean Square Error	2.557668

In our pursuit of predicting job performance, we meticulously followed a step-by-step process of collecting and refining the raw dataset. We began by sourcing and curating data from diverse sources, ensuring its comprehensiveness and accuracy. The dataset underwent rigorous cleaning, addressing missing values and outliers, thereby enhancing data quality.

Subsequently, we employed a predictive model to forecast individual job performance. Through iterative testing and refinement, we achieved notable results. Our model demonstrated an impressive accuracy rate of 83%, signifying its proficiency in making accurate predictions.

Further analysis revealed additional performance metrics Table 2. The F1 Score, with a micro value of 0.54 and a macro value of 0.1, illustrates our model's effectiveness in balancing precision and recall. Speaking of recall, it achieved a macro score of 0.34 and a micro score of 0.54, indicating its capacity to identify true positives effectively.

Our precision scores were also commendable, with a micro score of 0.54 and a macro score of 0.30, underscoring the model's precision in classifying positive instances. Additionally, our MAE (Mean Absolute Error) of 1.125, MSE (Mean Squared Error) of 6.5412, and RMSE (Root Mean Squared Error) of 2.5577668 collectively reinforce the model's predictive accuracy.

These comprehensive evaluation metrics underscore our commitment to enhancing predictive accuracy in assessing individual job performance. Our model's impressive results offer valuable insights, facilitating informed decision-making in talent management and human resource practices.

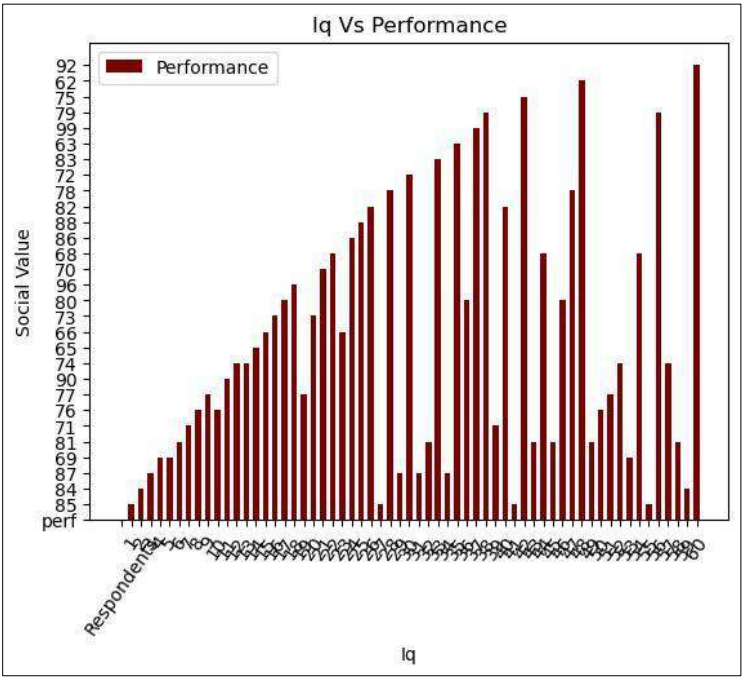


Figure 1: IQ vs performance

The bar graph visually illustrates the connection between IQ and job performance within our predictive model. Each bar represents the varying levels of IQ and their corresponding impact on performance in Figure 1. This graphical representation offers a clear and concise means to discern how IQ influences job performance, aiding in the model's interpretation. It serves as a valuable tool to comprehend the nuanced relationship between cognitive ability and professional success, enabling more informed decision-making in talent management and human resource practices.

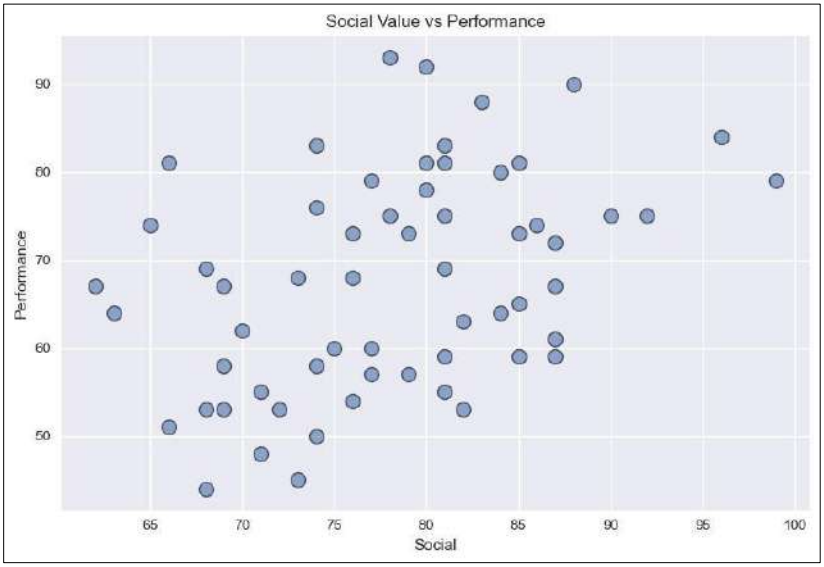


Figure 2: Social value vs performance

The scatter graph visually encapsulates the relationship between Social Value and Performance in Figure 2 within a job performance prediction model. This graphical representation serves as a powerful tool for intuitively comprehending the correlation between these two critical variables. By plotting individual data points, it unveils patterns, trends, and potential associations, providing a succinct and insightful view of how social value influences job performance. This visualization aids in the model's interpretability and assists decision-makers in understanding the impact of social attributes on overall job performance, ultimately contributing to more informed talent management strategies.

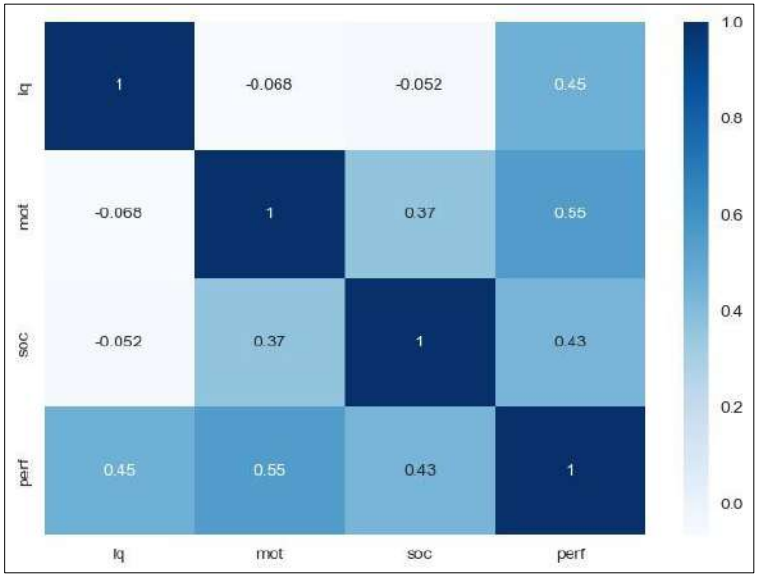


Figure 3: Heatmap representation

The heatmap in Figure 3 serves as a graphical representation illuminating the intricate interplay between critical factors in a job performance prediction model: Social Value, IQ, Performance, and Motivation (MOT). This visual representation offers a clear and concise insight into the multifaceted relationships among these variables. Darker regions signify stronger positive correlations, while lighter areas indicate weaker or negative associations. By leveraging the heatmap, we gain a deeper understanding of how Social Value, IQ, Performance, and Motivation collectively influence job performance. It acts as a valuable tool, aiding in the interpretation of complex relationships and contributing to more informed decision-making in talent management and performance optimization.

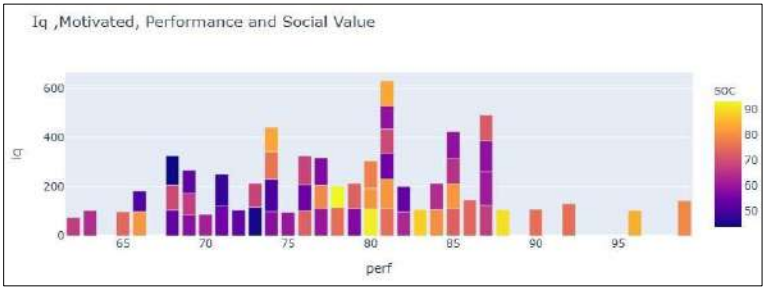


Figure 4: IQ vs motivated vs social value vs performance

The combined bar graph in Figure 4 offers a visual depiction of the intricate relationship between Social Value, IQ, Performance, and Motivation within a job performance prediction model. This graphical representation succinctly illustrates the correlations and interactions among these crucial factors. Through distinct bars for each variable, it provides a clear and concise view of how social value, cognitive aptitude (IQ), individual performance, and motivational factors intertwine in the context of predicting job performance. Such visual insights are instrumental in comprehending the multifaceted dynamics that shape an individual's professional success, aiding in informed decision-making and talent optimization.

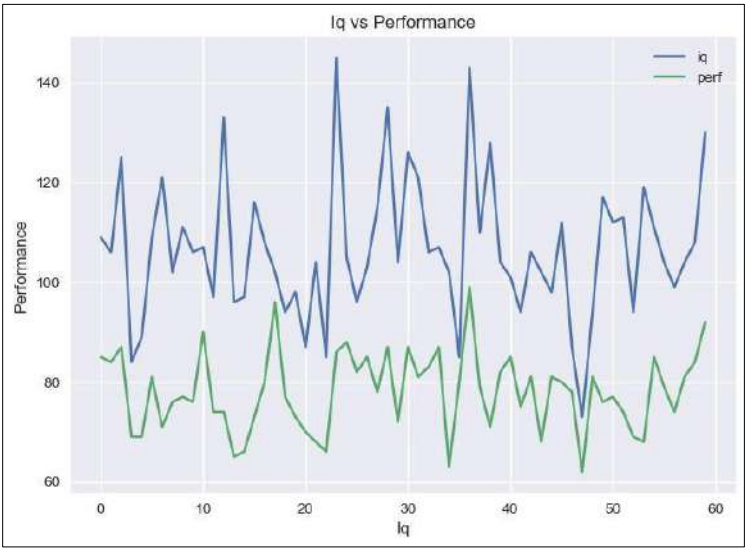


Figure 5: LogLoss

LogLoss in Figure 5, or Logarithmic Loss, serves as a critical gauge of a predictive model's overall performance in the context of a job performance prediction model, where variables like Social Value, IQ, Performance, and Motivations are at play. This metric quantifies the accuracy of predicted probabilities compared to actual outcomes. A lower LogLoss value indicates better model performance, signifying that the model effectively captures the complex interplay between these variables. By encompassing a holistic view of job performance predictors, LogLoss offers valuable insights, enabling organizations to make informed decisions in talent management and optimize their workforce with greater precision and confidence.

Conclusion

In conclusion, our journey in developing a predictive model for job performance assessment, considering critical variables such as Social Value, IQ, Performance, and Motivations, has been marked by a meticulous and data-driven approach. This endeavor involved multiple phases of data processing, culminating in a model that exhibits notable performance metrics.

Our model has demonstrated a commendable accuracy rate of 83%, indicating its ability to make correct predictions regarding job performance. This accomplishment is further underscored by the F1 Score of 0.54, reflecting a balance between precision and recall in classifying individuals.

Additionally, metrics such as MAE and MSE have provided insights into prediction errors and the model's overall performance. These scores serve as valuable indicators of the model's ability to accurately forecast job performance based on the selected variables.

It is important to note that while our model has yielded promising results, there remains room for further improvement. The performance metrics, though commendable, signify that there is still an opportunity to enhance predictive accuracy. This improvement can be achieved through more extensive and refined data collection processes, allowing the model to better capture the nuances of job performance determinants.

In summary, our predictive model represents a significant step towards informed talent management and human resource practices. While it has already proven its worth with an 83% accuracy rate and a F1 Score of 0.54, the pursuit of greater precision continues. Through continuous refinement and the collection of richer data, we aim to unlock even greater potential in accurately predicting job performance, ultimately contributing to more

effective talent optimization strategies and informed decision-making in the realm of human resources.

References

1. Decision tree methods: applications for classification and prediction,Yan-yan SONG and Ying LU, 2015 Apr 25; 27(2):doi: 10.11919/j.issn.1002-0829.215044
2. A study of job involvement prediction using machine learning technique, Youngkeun Choi, Jae Won Choi,doi 7 May 2021
3. T. M. B, Intelligent Human Centered Computing, vol. 1. Singapore: Springer Nature Singapore, 2023. doi: 10.1007/978-981-99-3478-2.
4. X. Fang, M. Xu, S. Xu, and P. Zhao, "A deep learning framework for predicting cyber attacks rates," Eurasip J. Inf. Secur., vol. 2019, no. 1, 2019, doi: 10.1186/s13635-0190090-6.
5. S. Mehtab and J. Sen, "A Robust Predictive Model for Stock Price Prediction Using Deep Learning and Natural Language Processing," SSRN Electron. J., 2020, doi: 10.2139/ssrn.3502624.
6. R. Vinayakumar, K. P. Soman, and P. Poornachandran, "Evaluating deep learning approaches to characterize and classify malicious URL's," J. Intell. Fuzzy Syst., vol. 34, no. 3, pp. 1333-1343, 2018, doi: 10.3233/JIFS-169429.
7. Y. Xu, Y. Zhou, P. Sekula, and L. Ding, "Machine learning in construction: From shallow to deep learning," Dev. Built Environ., vol. 6, no. February, p. 100045, 2021, doi: 10.1016/j.dibe.2021.100045.
8. S. Mandal, R. Sarkar, and S. Sinha, "Mathematical models of malaria - A review," Malar. J., vol. 10, pp. 1-19, 2011, doi: 10.1186/1475-2875-10-202.
9. S. Bhat, S. Bhat, R. Raju, R. D'Souza, and K. G. Binu, "Collaborative learning for outcome based engineering education: A lean thinking approach," Procedia Comput. Sci., vol. 172, no. 2019, pp. 927-936, 2020, doi: 10.1016/j.procs.2020.05.134.
10. Job Satisfaction Prediction and Machine Learning Technique, Youngkeun Cho, Jae Choi, doi.org/10.21203/rs.3.rs-1683972/v1
11. Lather, A.S.; Malhotra, R.; Saloni, P.; Singh, P.; Mittal, S. Prediction of Employee Performance Using Machine Learning Techniques. In

Proceedings of the International Conference on Advanced Information Science and System, Singapore, 15- 17 November 2019;

12. Obiedat, R.; Toubasi, S.A. A Combined Approach for Predicting Employees' Productivity based on Ensemble Machine Learning Methods. Informatica 2022, 46, 1.

Chapter - 24

Model for Predicting Electricity Prices Using Chaos Theory

Authors

Rituparna Mitra
Electrical Engineering, Swami Vivekananda University, Kolkata, West Bengal, India

Avik Datta
Electrical Engineering, Swami Vivekananda University, Kolkata, West Bengal, India

Promit Kumar Saha
Electrical Engineering, Swami Vivekananda University, Kolkata, West Bengal, India

Titas Kumar Nag
Electrical Engineering, Swami Vivekananda University, Kolkata, West Bengal, India

Susmita Dhar Mukherjee
Electrical Engineering, Swami Vivekananda University, Kolkata, West Bengal, India

Suvraujjal Dutta
Electrical Engineering, Swami Vivekananda University, Kolkata, West Bengal, India

Chapter - 24

Model for Predicting Electricity Prices Using Chaos Theory

Rituparna Mitra, Avik Datta, Promit Kumar Saha, Titas Kumar Nag, Susmita Dhar Mukherjee and Suvraujjal Dutta

Abstract

The forecasting model for electricity prices presented in this research is based on chaos theory. The chaotic aspect of energy pricing is first confirmed using the chaos theory. The attractors' fractal dimensions and Lyapunov exponents are retrieved. Here, it is evident that the power price exhibits chaotic properties, giving the foundation for carrying out a short-term electricity price projection using the chaos theory. Then, using multivariable time series made up of the electricity price and its linked parameters, such as the time series for the system load and the available generating capacity, an accurate phase space is rebuilt. The global and local energy price forecasting models based on the recurrent neural network are constructed, and with these models, the electricity prices in the New England power market are successfully forecasted. This is done by tracking the evolving trend of the neighboring phase points in the phase space.

Keywords: Chaos, electricity price, forecast, power market

1. Introduction

The global electric power sector has entered a new phase of market operation and deregulation. The price of electricity plays a significant role as an economic lever in the power market, where it is exchanged as a commodity. The evolution of the price of electricity embodies a diversity of market information that fully reflects the laws of motion of many market elements. Analysis and understanding of the law of evolution of the price of electricity will assist market administrators in creating efficient market regulations as well as market participants in setting up logical trading strategies to maximise their financial gain.

The market's supply and demand, as well as the transmission system's

operational circumstances, all have an impact on the price of power. These elements include the weather, the state of the economy, development planning, and failures and accidents. These components work together to produce the incredibly complex random variance in electricity prices. The energy price is recognized as a stochastic variable because of its extremely complex variance, which is why its development process is typically considered to be stochastic behavior. Now, it is customary to examine its distribution function and create its forecasting model using the theory of probability statistics [1].

However, it has recently been discovered that many equally complex systems, such as the stock price, traffic flow, and power load, actually have definite regularity and order hidden behind their seemingly random appearance; in other words, their evolving processes appear chaotic rather than the commonly believed completely random behavior.

Additionally, the chaos theory is used to thoroughly analyses and forecast their changing processes. It is not accurate enough to investigate the electricity price as a stochastic variable just with the science of probability and statistics if the seemingly random evolution of the energy price also possesses the chaotic quality. Instead, until more precise and logical analysis results and prediction models are obtained, the chaos theory can be employed to highlight its inherent regularity.

Based on the foregoing study, the chaos theory is used to first confirm the intrinsic law of what appears to be a random changing process of the chaotic feature of the power price.

Second, a multi-variable time series made up of the electricity price and its correlated elements, such as the time series for system load and available generating capacity, is used to rebuild an accurate phase space in order to do accurate short-term forecasting. The recurrent neural network is used to track the changing trend of the nearby phase points in the phase space, accurately building the electricity price forecasting model.

i) Chaotic characteristics of electricity price

a) Reconstruction of phase space

The price of electricity is influenced by a number of interrelated factors, including the architecture of the transmission network, customer and power company bid strategies, the environment and the state of the economy. A very complex and sizable dynamic system is created by the power price and

its connected factors, which change over time and display very complex characteristics. The phase space, which is constructed using all state variables as coordinates, is an efficient and understandable tool for observing the evolution orbits of the state variables and further exploring the underlying law of the dynamic system if a nonlinear system can be described by a set of differential equations formed by the state variables. It is, however, impossible to create its explicit mathematical differential equations and unrealistic to acquire the actual phase space of the electricity price system due to the complexity of the dynamic system of the electricity price.

F. Takens successfully applied the differential topological technique in 1980 to demonstrate that a single variable can rebuild the attractors of a non-decoupled system, which are the invariant manifolds or point set of the ultimate evolutionary trend of a non-linear system in phase space [2].

As a result, a phase space that preserves the topological characteristics of the original attractors of the electricity pricing system can be recreated using the single variable of the electricity price. The evolution characteristics of the price of power can be examined using this phase space that has been created. Fig. 1 displays the hourly power prices (totaling 4368 points) for the New England market in the United States from December 1, 2001 to May 31, 2002 (data obtained from the website "http://www.iso-ne.com"). This pricing curve behaves in a very erratic manner. Assume that we have the observed data for the price of electricity contained in the m-dimensional phase space as a single variable time series, x1, x2,…, and x_N. The following is an expression for a sequence of phase points in the m-dimensional phase space:

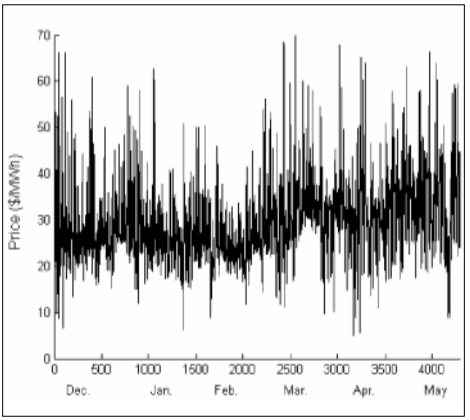


Figure 1: Price curve

The embedding dimension m and the delay period must be carefully chosen in order to rebuild the phase space. The mutual information approach, which establishes the delay time when the mutual information first reaches minimum, is advised as the best option [3]. Kennel *et al*. [4] presented the fake closest neighbor approach to get the minimum necessary embedding dimension m.

Due to the very low dimension of the space, the false neighbors are phase points that are quite close to one another. Less false neighbors will exist as the embedding dimension is increased, allowing for the selection of an acceptable embedding dimension.

In Fig. 2, with the delay time τ=8, the mutual information is estimated for the time series of the electricity price in the New England market.

The relationship between the fractions of false neighbors and the embedding dimension is depicted in Fig. 3. The false neighbors equal 0, which proves that the embedding dimension is 9, as can be shown from Fig. 3 if the embedding dimension m=9.

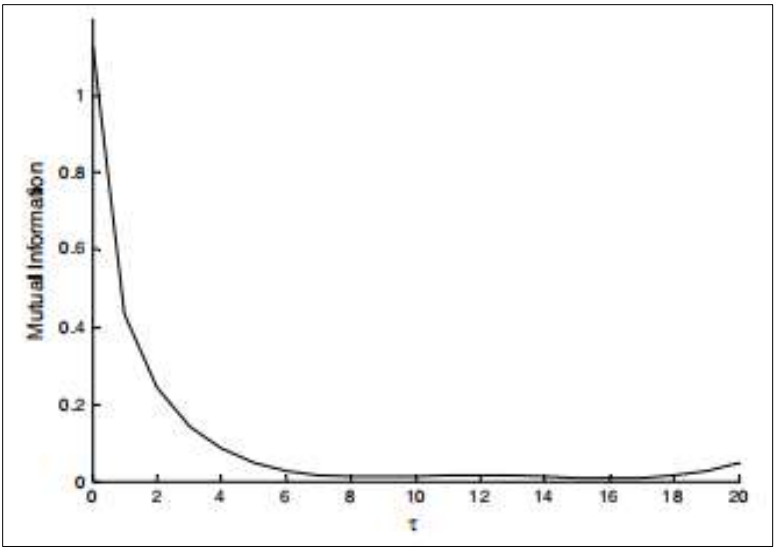


Figure 2: Manual Information curve

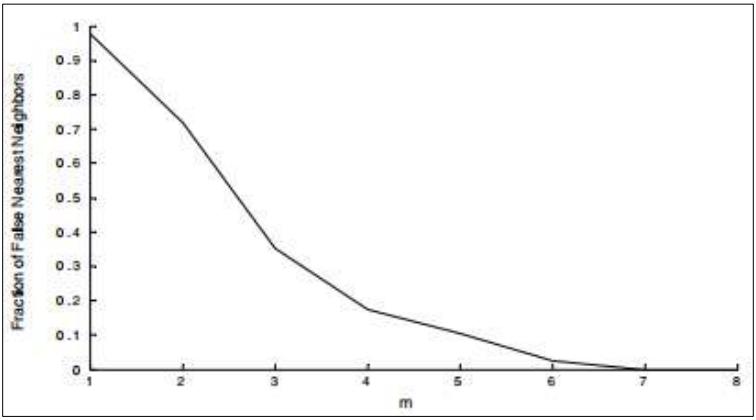


Figure 3: Curve of false neighbors and embedding dimension

b) An analysis of the chaotic characteristics

The main characteristics of chaos are its sensitive dependence on the initial conditions and its self-similar fractal structure with a non-integral dimension, or fractal dimension. Chaos is an irregular, seemingly random behavior that occurs in a deterministic system without the assistance of additional random factors. It is possible to extract several significant characteristics of the attractors, such as the fractal dimension and Lyapunov exponent, in order to determine if the evolution of the power price is chaotic or not.

1) Fractal dimension: In 1983, Grassberger and Procaccia put forth the G-P technique, which involves using the correlation integral C(m,r) to extract the correlation dimension D2 of attractors in the reconstructed phase space [5].

The relationship curve C(m, r) =ln r. ln ln for the dynamic system of the electricity price in the New England market is shown in Fig. 4 and Table 1 using the G-P method. So, by figuring out the slope of the straight line of the curve, it is possible to determine the correlation dimension D2 (m). As the embedding dimension is increased, the correlation dimension D2 (m) eventually approaches constant, as seen in Fig. 4.

The correlation dimension D2 (m), which is practically constant if the embedding dimension is 7 or above, and D2 = 4.5, which is not an integer, show that the power price's evolution in the New England market is chaotic.

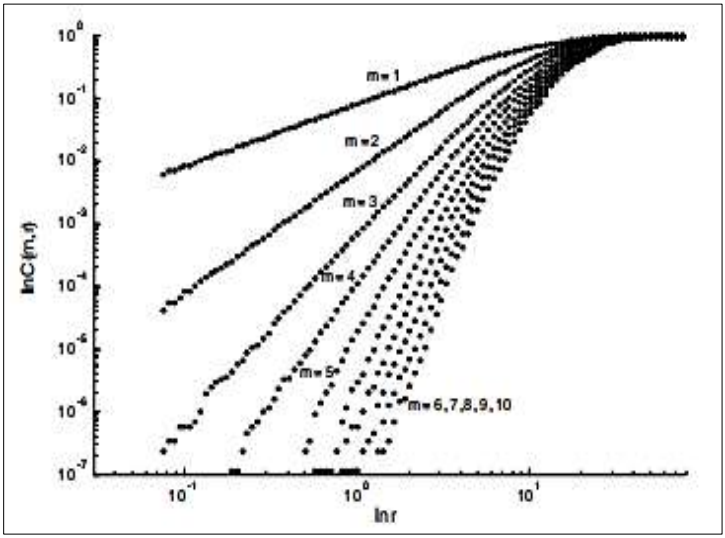


Figure 4: In$C(m,r)$~ In r curve

Table 1: Correlation D dimensions for different embedding dimensions

Embedding dimension	6	7	8	9	10
Correlation dimension	5.31	5.39	5.42	5.45	5.48

2) Lyapunov exponent: The Lyapunov exponent is a measure for quantitatively describing this feature of chaos and a criterion for determining whether a dynamic system is chaotic or not. If the Lyapunov exponent is greater than 0, the dynamic system is chaotic. The Lyapunov exponents are extracted from the phase space using the small data algorithm [6], and are displayed in Table 2. The basic characteristic of the chaotic system is its sensitive dependence on the initial conditions. This demonstrates how highly sensitive the price of power is to its starting point. However, because the Lyapunov exponents are so small, the rate of evolution of the price of electricity is quite modest, demonstrating that accurate short-term forecasting of the power price is conceivable.

Table 2: Lyapunov exponents

Embedding dimension	6	7	8	9	10
Lyapunov exponent	0.020	0.016	0.014	0.013	0.012

ii) Forecasting model of electricity price

a) Reconstruction of Phase Space from multivariable time Series

Typically, the evolving trend of the phase points in the phase space, which is rebuilt by the single time series, is used to establish the forecasting model for chaotic time series. If only rational embedding dimension and delay time are selected, the single time series are typically sufficient to reconstruct the phase space of the dynamic system, according to F. Taken's embedding theory, and the forecasting model thus developed can produce fairly satisfactory forecasting results. The time series that are obtained in actual issues, however, are frequently noisy and have length restrictions. As a result, the phase space generated by a single time series cannot adequately capture the changing trends of the dynamic system's state variables. Furthermore, it is not difficult to obtain multivariable time series in many real-world circumstances, which can contain more data and reconstruct a more precise phase space, further enhancing the forecasting model's accuracy [7].

The system load, the time series of available generating capacity, and many other elements all work together to determine the price of energy. When the electricity market is actually in operation, time series for the system load and available generating capacity can also be found, which together make up the multivariable time series.

b) Global and local electricity price forecasting model

Based on RNN

The key to developing the forecasting model will be understanding how to solve (2) and (3) to arrive at the ideal function while establishing the forecasting function. The neural network's potent mapping capability has recently come to light thanks to the advancement and application of neural network theory. Due of its capacity for dynamic mapping, the recurrent neural network (RNN) is preferred among them.

The RNN has demonstrated its dynamic properties through the calculation process, demonstrating not only its capacity to map the complex function but also its superiority in time convergence, making it appropriate for creating the chaotic system forecasting model.

The predicting function is mapped using the Elman model, one of the RNN models, in this study. The input unit, the hidden unit, the output unit,

and the recurrent unit which is utilised to remember the hidden units' earlier activation are the four components that make up the Elman network. There are two steps in the Elman neural network's training process. The global approach is used in step 1, meaning the neural network is trained using samples created from every phase point, completely charting the evolution of every phase point. This stage typically takes place offline and has a lengthy calculation cycle. In step 2, the local approach is applied, which entails retraining the trained neural network using samples constructed from the closest phase points of the present state in order to closely follow the evolution of the most recent phase points. This step, whose calculation cycle is the forecasting step T, is typically realized online.

c) Forecasting results analyzed

Two time series, namely the electricity price time series and the system load time series, are gathered in the New England market. The two-time series are used to reconstruct the phase space. For the electricity pricing time series, the embedding dimension is 9 and the delay time is 8, while for the system load time series, the embedding dimension is 8 and the delay time is 5. For the forecast, the electricity costs for the dates of March 27 and April 18 are chosen. On March 27, the price of power changed just slightly, whereas on April 18, the price of electricity changed significantly. The lengths of the forecasting steps are assumed to be 1, 25, and 49, corresponding to forecasting one hour, one day, and two days in advance, respectively. The forecasting results are shown in Fig. 5 and the forecasting errors are shown in Table 3. The forecasting errors of the price of power are acceptable, as shown by Fig. 5 and Table 3, demonstrating the viability of a short-term forecast of the price of electricity and the efficacy of global and local RNN-based forecasting models.

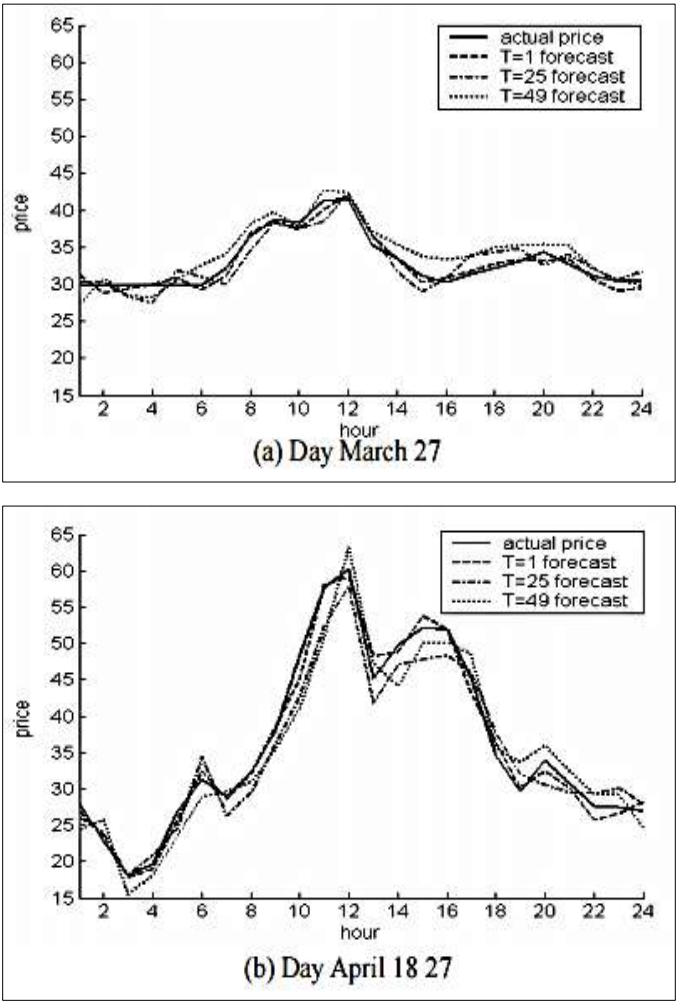


Figure 5: Price forecastine curve

Table 3: Forecastine eorrors

Step (h)	Forecasting errors on day March 27		Forecasting errors on day April 18	
	Mean (%)	Max. (%)	Mean (%)	Max. (%)
1	2.22	4.66	3.13	4.66
25	4.42	7.90	6.80	11.28
49	5.09	9.78	8.00	14.76

Conclusions

This research examines the time series of power price in the New England electrical market using the notion of chaos. It has been discovered that chaos best describes the way in which the price of power is changing. The global and local power price forecasting models based on the RNN are proposed using the results and by tracking the evolving trend of the nearby phase points in the phase space. The price of electricity in the New England market is accurately predicted using the suggested forecasting models.

References

1. F. J. Nogales, A. J. Conejo, R. Espinola, "Forecasting next-day electricity prices by time series models", IEEE Trans. on Power System, Vol. 17, pp. 342-348, May, 2002.
2. T. S. Parker, L O. Chua, Practical numerical algorithms for chaotic systems, Berlin: Springer Verlag, 1989.
3. A. M. Fraser, H. L. Swinney, "Independent coordinates for strange attractors from mutual information", Physical Review A, Vol. 33, pp. 1134-1140, 1986.
4. M. B. Kennel, R. Brown, H. D. I. Abarbanel, "Determining embedding dimension for phase-space reconstruction using a geometrical construction", Physical Review A, Vol. 45, pp.3403-3411, 1992.
5. P. Grassberger, I..Procaccia, "Measuring the strangeness of strange attractor", Physica D, Vol. 9, pp.198-208, 1983.
6. M. T. Rosenstein, J. J. Collins, C. J. De Luca, "A practical method for calculating largest Lyapunov exponents from small data sets", Physica D, Vol. 65, pp.117-134, 1993.
7. C. Liangyue, M. Alistair, J. Kevin, "Dynamics from multivariate time series", Physica D, Vol. 121, pp.75-88, 1998.

Chapter - 25

Electric Mobility's Cost-Optimal Market Share in the Energy System under a Decarbonization Scenario

Authors

Rituparna Mitra
Electrical Engineering, Swami Vivekananda University, Kolkata, West Bengal, India

Avik Datta
Electrical Engineering, Swami Vivekananda University, Kolkata, West Bengal, India

Rituparna Mukherjee
Electrical Engineering, Swami Vivekananda University, Kolkata, West Bengal, India

Promit Kumar Saha
Electrical Engineering, Swami Vivekananda University, Kolkata, West Bengal, India

Titas Kumar Nag
Electrical Engineering, Swami Vivekananda University, Kolkata, West Bengal, India

Suvraujjal Dutta
Electrical Engineering, Swami Vivekananda University, Kolkata, West Bengal, India

Chapter - 25

Electric Mobility's Cost-Optimal Market Share in the Energy System under a Decarbonization Scenario

Rituparna Mitra, Avik Datta, Rituparna Mukherjee, Promit Kumar Saha, Titas Kumar Nag and Suvraujjal Dutta

Abstract

By 2050, the EU wants to have cut greenhouse gas emissions by at least 80%. While emissions in the electricity and heating sectors have decreased over the past few years, this has not been the case for the transportation sector. Particularly electric vehicles that are powered by renewable energy sources have a significant potential to cut greenhouse gas emissions. Their increased demand for power has a significant impact on the power system if their population grows significantly. The energy system model, SCOPE, models long-term decarbonization scenarios and optimizes investment choices for the transportation sector endogenously. It encompasses the power, heat, and transportation sectors. We examine the ideal distribution of various vehicle types and contrast several situations with regard to the distribution of controlled-charging cars and the accessibility of e-highway trucks. Even with long-term low power to liquid import prices, electric mobility has obvious economic advantages. EHighway, trucks are cost-efficient in all scenarios.

Keywords: Electric vehicles, energy system model, optimization, sector coupling, ehighway truck.

1. Introduction

Until 2050, the EU wants to cut greenhouse gas emissions by at least 80%. It will take the combined efforts of all sectors to achieve this aim. In Germany, the transportation sector is responsible for around 18% of greenhouse gas (GHG) emissions and nearly 33% of total energy consumption. While emissions in the electricity and heating sectors have decreased over the past few years, this has not been the case for the transportation sector. Particularly electric vehicles provide significant

potential to cut GHG emissions in the transportation sector when powered by renewable energy sources. If the number of electric vehicles significantly rises, the additional power they use has a huge impact on the power supply. Therefore, in long-term decarbonization scenarios, the power demand of electric vehicles and its impact on the power infrastructure must be taken into account together. We outline a potential modelling strategy for the integrated optimization of the power, heat, and transportation sectors within the "SCOPE" energy system model that endogenously optimizes investment and dispatch choices in a long-term decarbonization scenario. Additionally, we distinguish between flexible-charged cars and unregulated-charged vehicles and simulate both of them simultaneously. While [1] offers a framework for representing the road transportation sector in energy system models and evaluates scenarios for the UK with a focus on 2050, [2] describes an applied modelling approach with a focus on Germany until 2030 [3] also emphasizes Germany.

The rest of this piece is organized as follows: Section II describes the applicable model, while Section III provides a summary of the assumptions. Section IV presents the results of the model. The conclusion brings the essay to a close.

2. Model description

a) Overview

The analysis is conducted using the "SCOPE" cross-sector expansion and dispatch optimization model, which establishes a low-cost energy supply system (for the electricity, heat, and transportation sectors) [4]. The target level of GHG reduction might be stated. It is guaranteed that the supply will be secure or that the demand will always be met in all applications and industries with hourly resolution. It is a deterministic, linear model that was created as part of the "Interaction of Electricity, Heat, and Transport" project [5] and is based on a variety of earlier studies (see [6], [7]). Thus, the economic optimization can determine both the most cost-effective power plant dispatch and the wisest investment choices based on flawless foresight. It is distancing itself from the current legal system. Europe, excluding the Balkans and the Baltic States, is the geographic scope of the energy system. The model is shown schematically in Figure 1. In addition to the markets for electricity and gas, there is also a general market for emission permits. The heat and transportation sectors are also fully depicted. Each of the provided technologies engages with several markets. The potentials for wind energy

and photovoltaics (PV), which result from area constraints and distance regulations, are crucial input data for the model. High-resolution time series for the feed-in of PV and wind energy are also included, both temporally and geographically. The moving water, stored water, and pumped storage power plants' historical inflow data are used to model the European hydrological power plant park [8].

b) Transport sector

Average driving profiles are required for the simulation of the various passenger and freight transport choices, including market segmentation and vehicle sizes. As a result, each transport market segment's hourly driving profile time series Dtotal are produced using a multi-stage stochastic technique. To achieve this, thorough analyses of German mobility patterns are conducted in order to create a strong data base for plausible parameters and empirical distributions. Additionally, a vehicle fleet model is used to light commercial vehicles (lcv) and passenger automobiles in order to establish realistic upper and lower boundaries for the various drive concepts and market categories. The model combines a bottom-up consumer demand method with a dynamic stock-flow model, allowing for a quantitative determination of the future vehicle fleet under various general conditions (such as fuel prices, technological costs, or regulatory framework) (for more information, see [9], [10]). The energy system model "SCOPE" incorporates the aforementioned time series as well as upper and lower bounds for the various technologies. The model distinguishes between four electric engine driving concepts and internal combustion engine (ICE) cars (gasoline, diesel, and natural gas). These include hydrogen (H2) vehicles, range-extender electric vehicles (REEV), battery electric vehicles (BEV), and plug-in hybrid vehicles (PHEV). Additionally, a number of transport market categories are distinct. The size of vehicles is divided into small, medium, large, lcv, and truck categories on the one hand. The annual distance, however, is divided into three categories. Only trucks can be distinguished by size. The charging of the cars can be controlled or uncontrolled. Models are also used to examine the possibility of returning battery energy to the power grid (vehicle-to-grid, or V2G).

The traditional power demand coverage equation includes the energy needs of the transportation sector, but the GHG emission constraint also includes emissions from fossil fuels. The target function takes all vehicles' fixed and variable expenses into account. Infrastructure costs (such as those

associated with charging, overhead wires, and H2-related infrastructure) are included in the variable vehicle costs but are not explicitly modelled.

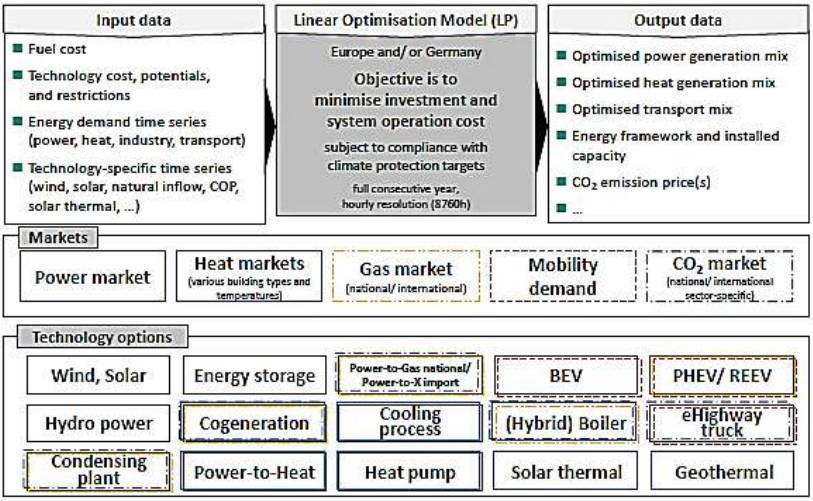


Figure 1: Overview of the model SCOPE for investment and dispatch optimization across energy sectors

$$\mathbf{D^{el}_{t,v,s}} + \mathbf{D^{ICE}_{t,v,s}} = D^{total}_{t,s} \cdot \mathbf{x_{v,s}} \tag{1}$$

$$\lambda_{v,s} \cdot D^{el,max}_{t,v,s} \cdot \mathbf{x_{v,s}} \leq \mathbf{D^{el}_{t,v,s}} \leq D^{el,max}_{t,v,s} \cdot \mathbf{x_{v,s}} \tag{2}$$

$$\begin{aligned} &\lambda_{v,s} \cdot (D^{total}_{t,s} - D^{el,max}_{t,v,s}) \cdot \mathbf{x_{v,s}} \leq \mathbf{D^{ICE}_{t,v,s}} \\ \leq &(D^{total}_{t,s} - \lambda_{v,s} \cdot D^{el,max}_{t,v,s}) \cdot \mathbf{x_{v,s}} \end{aligned} \tag{3}$$

$$\begin{aligned} &(1-\lambda_{v,s}) \cdot SOC^{min}_{t,v,s} \cdot \mathbf{x_{v,s}} \leq \mathbf{SOC_{t,v,s}} \\ &\leq (1-\lambda_{v,s}) \cdot SOC^{max}_{t,v,s} \cdot \mathbf{x_{v,s}} \end{aligned} \tag{4}$$

$$\begin{aligned} \mathbf{SOC_{t+1,v,s}} &= \mathbf{SOC_{t,v,s}} - \mathbf{D^{el}_{t,v,s}} \cdot \eta^{el}_{v} \\ &+ \lambda_{v,s} \cdot \mathbf{x_{v,s}} \cdot D^{el,max}_{t,v,s} \cdot \eta^{el}_{v} \\ &+ \mathbf{P^{charge}_{t,v,s}} \cdot \eta^{charge}_{v} \cdot 1h \\ &- \lambda_{v,s} \cdot \mathbf{x_{v,s}} \cdot P^{charge,inflex,fix}_{t,v,s} \cdot \eta^{charge}_{v} \cdot 1h \\ &- \mathbf{P^{discharge}_{t,v,s}} / \eta^{discharge}_{v} \cdot 1h \end{aligned} \tag{5}$$

$$\begin{aligned} &\lambda_{v,s} \cdot P^{charge,inflex,fix}_{t,v,s} \cdot \mathbf{x_{v,s}} \leq \mathbf{P^{charge}_{t,v,s}} \\ &\leq \lambda_{v,s} \cdot P^{charge,inflex,fix}_{t,v,s} \cdot \mathbf{x_{v,s}} \\ &+ (1-\lambda_{v,s}) \cdot P^{charge,flex,max}_{t,v,s} \cdot \mathbf{x_{v,s}} \end{aligned} \tag{6}$$

$$0 \leq \mathbf{P^{discharge}_{t,v,s}} \leq (1-\lambda_{v,s}) \cdot P^{charge,flex,max}_{t,v,s} \cdot \mathbf{x_{v,s}} \tag{7}$$

$$x^{min}_{v,s} \leq \mathbf{x_{v,s}} \leq x^{max}_{v,s} \tag{8}$$

$$\sum_{v} \mathbf{x_{v,s}} = N_s \tag{9}$$

3. Model assumptions

In an ambitious scenario for the year 2050 where the objective for greenhouse gas emission reduction is set at -95% (base year 1990) in compliance with the Paris Climate Agreement, we wish to examine which vehicle options are most suited for a cost-effective decarbonization. All sources of greenhouse gas emissions, both energetic and non-energetic, are considered for this balance, with the exception of the influence of international ship and aviation traffic. This results in a zero-emission target for the total amount of energetic emissions because the decrease of non-energetic emissions is constrained. Additionally, to this overarching objective, SCOPE requires other input data. Hourly time series for specific electricity and heat demands, the availability of renewable electricity generation that is weather dependent, as well as the demand for transportation in road transportation, are all included in the input data. Furthermore, exact extension limits for the modeled units must be defined together with technical specifications, investment and operating expenses, and other factors. We consider the feasibility of importing hydrogen (113 Euro/MWhfuel) as well as power-based liquid fuels from outside of Europe at a cost of 118 Euro/MWhfuel [11]. These expenses do not include grid fees or charges for the country's infrastructure. The comprehensive parameter set that served as the foundation for the findings displayed in the following section is provided in [12]. Four main scenario options for road transportation are examined: For the baseline scenario, we assume that BEV and PHEV/REEV share flexible charging at rates of 80% and 60%, respectively. Additionally, we presumptively have the infrastructure for eHighway trucks. under comparison, we lower the proportion of vehicles with flexible charging to 60% (BEV) and 20% (PHEV and REEV) under the "reduced flexibility" scenario. The infrastructure for eHighway trucks is assumed to be unavailable in the "without eHighway trucks" scenario, which eliminates the prospect of investing in eHighway trucks. Finally, we assume the ability to feed electricity back to the grid in the "V2G" (vehicle-to-grid) scenario. Between the several scenario options, the underlying assumptions for the remaining energy system are unchanged.

4. Results

The primary electricity producing technology in the majority of nations are PV and wind power plants, both onshore and offshore. In order for Europe to meet the climate objective, 1,210 GW of solar PV, 803 GW of onshore wind, and 143 GW of offshore wind facilities must be developed. Other renewable energy sources including hydropower, geothermal power,

and biomass-based power plants only contribute a small part. Additionally, fuel-based power plants are used to provide system security and for the combined generation of heat and electricity (combined heat and power plants, combined cycle gas turbines, and gas turbines, as well as the few existing nuclear power plants). Existing pump turbines (63 GW) and internally generated batteries (156 GW) complete the system in European System. Germany's electricity output and consumption are shown in Fig. 2. Germany's annual energy consumption is over 800 TWh. The 450 TWh of conventional consumption and grid losses, 150 TWh of heating and cooling needs, 180 TWh of direct usage in the transportation sector, and 25 TWh of power to gas (P2G) generation make up the total energy demand. 195 TWh of climate-neutral fuels are needed for national transportation, as well as for the production of fuel-based power and heat, in order to meet the climate target. To create the necessary power to liquid (P2L) or P2G fuels (often PtX), this results in an additional electricity demand of about 400 TWh outside of Europe or at specialized offshore stations. The full findings of cost-optimized market shares of various driving ideas for road transportation are shown in Fig. 3. The diagram is organized as follows: The market shares of the base case scenario are presented at the bottom (Fig. 3c). They are provided as a proportion of all cars for each sector that was modeled. We make a distinction between light and heavy commercial vehicles as well as small, medium, and big passenger cars. Once more, each of these magnitudes is broken down into three groups based on the normal yearly mileage of pcs and lcvs or the weight category for heavy commercial vehicles. Fig. 3a shows the total mileage of all matching vehicles in passenger kilometers (pkm) or ton kilometers (tkm), which are used to determine the importance of each segment in road transportation. The variations in market shares between the baseline scenario and the scenario with less flexibility are shown in Fig. 3b. According to the baseline scenario (Fig. 3c), electric mobility (BEV, REEV, and PHEV) is the top driving idea for a future involving zero-emission road transportation. Even with persistently low P2L import prices, vehicles dependent solely on fuel play a modest role. Only passenger cars with extremely low annual miles are utilized with them. The improved efficiency of the electric engine becomes more significant with larger annual mileage, which results in very high shares of all-electric cars in segments >14,000 km/a. PHEV (parallel hybrid drive with smaller battery) appears to have an edge over REEV (serial hybrid drive with larger battery and small combustion engine) in the majority of market segments due to the long-term cost degression for batteries in the passenger car area. Electric driving ideas are also dominant within the

realistic possible ranges for heavy commercial truck segments (BEV for trucks under 12 t; eHighway trucks and PHEV for semitrailers above 40 t, which is the main truck category in terms of total tkm). For the cost suppositions and feasible electric driving shares, electric driving choices are only not cost-effective for the truck segment of 12 to 28 t. The results of the baseline scenario are affected by the assumption of reduced charging flexibility, making BEVs less desirable for the entire energy system (see Fig. 3b). Gasoline vehicles partially replace BEVs for vehicles with low annual mileage, and REEVs partially replace BEVs for vehicles with high annual mileage. In terms of big commercial vehicles, hydrogen trucks that weigh up to 12 t replace a significant portion of BEVs, while hydrogen trucks that weigh more than 40 t replace a portion of PHEV. Only the segment of large commercial vehicles > 40 t, where eHighway trucks have a market share of 75% in the base case, would change if no eHighway trucks were available. The hydrogen trucks and PHEV would take over this market share around 50/50. Only a small amount of market shares for vehicles are impacted by the ability to feed back power from vehicles. BEVs can only join the market and partially replace PHEVs in the small passenger car class, where they now have a market share of about 25%. The various scenarios affect installed capacity of power generating and storage technologies in addition to market shares of specific vehicle types (Fig. 4). The demand for batteries and CCGTs rises in the scenario with less flexible charging, and more P2G plants and PV are required. The direct power demand in Germany would be lower without eHighway vehicles, necessitating less wind and PV installation. However, compared to the baseline scenario, the demand for fuel-based energy carriers—in particular renewable hydrogen—increases by around 40 TWh/a, which drives up overall electricity demand. Due to the relatively rigid power demand that eHighway vehicles represent, the flexibility demand is typically lower. The ability to feed electricity generated by moving vehicles back into the grid eliminates the need for gas turbines, CCGTs, and batteries. In its place, more solar and wind energy is used to power electric automobiles.

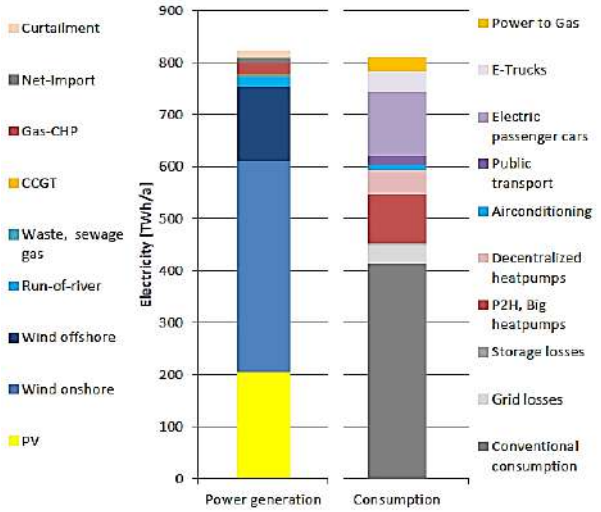


Figure 2: Electricity production and consumption in Germany in the baseline scenario for 2050

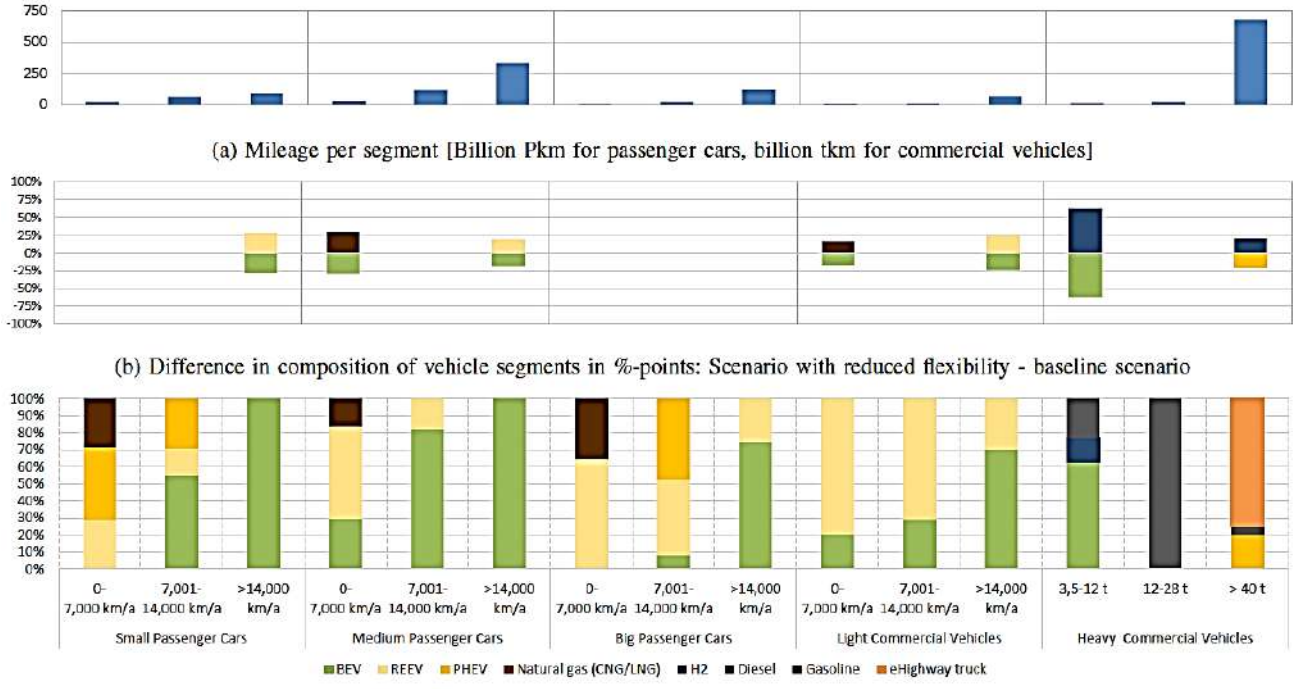


Figure 3: Model results of the investment in passenger cars and light as well as heavy commercial vehicles in the 2050 scenarios

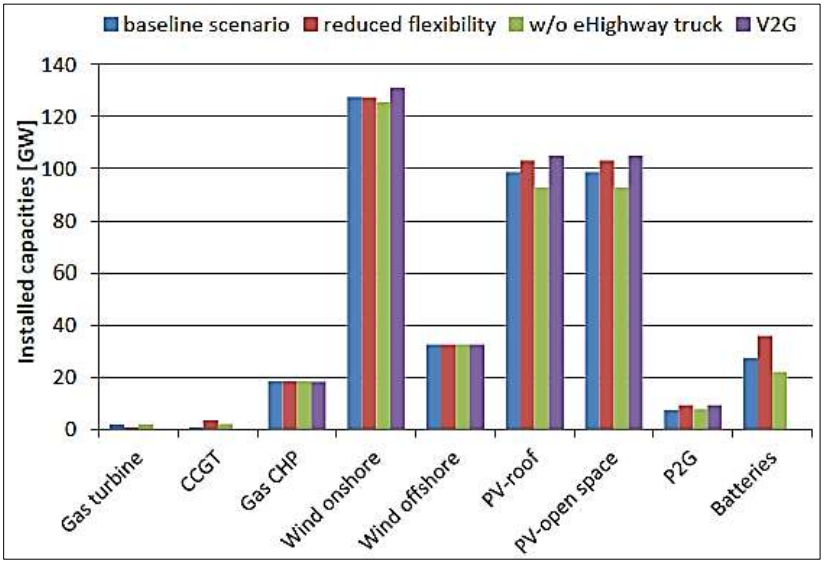


Figure 4: Installed capacities of power generation and storage technologies in Germany for the 2050 scenarios

5. Conclusion

In this article, we developed a technique for examining the transportation industry's potential contributions to a 2050 decarbonization scenario.

The majority of the nation's PtX need is supplied at extremely low cost by imports. Due to a momentarily extremely high electrical supply, there is also a very little national economic P2G potential that is competitive with PtX imports.

Utilizing electric mobility's potential greatly minimizes the requirement for P2L imports. The V2G option results in a decrease in the demand for stationary storage and backup power plants and increases the share of PV. Electric car flexibility restrictions have the reverse impact of V2G, though to a much lesser amount. Even with consistently low P2L import prices, electric mobility provides distinct cost advantages. REEV (serial hybrid drive with larger battery and tiny combustion engine) appears to have an advantage over PHEV (parallel hybrid drive with smaller battery) in all market categories due to the anticipated long-term cost degression for batteries in the passenger automobile field. Hybrid vehicles are more appealing than all-electric vehicles for drivers with low annual mileage. In

every case, BEV trucks in the field of 12 t and eHighway trucks in the major segment > 40 t are more economical. However, there are still questions about the costs and feasible electric driving shares in the world of trucks. The advantage of H2 trucks over P2L trucks would depend on H2 infrastructure expenses if overhead line infrastructure hadn't been installed. The cost difference between truck segments is typically small, so it is necessary to evaluate costs and emissions along the trip (from now to 2050).

References

1. P. E. Dodds and Will McDowall, "Methodologies for representing the road transport sector in energy system models", *International Journal of Hydrogen Energy*, vol. 39, no. 5, pp. 2345-2358, 2014.
2. H. U. Heinrichs, *Analyse der langfristigen Auswirkungen von Elektromo-bilität auf das deutsche Energiesystem im europäischen Energieverbund*, 2013, ISBN 978-3-7315-0131-2.
3. A. Palzer, Sekioriibergreifende Modellierung und Optimierung eines zukünftigen deutschen Energiesystems unter Berücksichtigung von Energieefjizienzmaβnahmen im Gebäudesektor, Fraunhofer Verlag, 2016, ISBN 978-3-8396-1041-1.
4. "Fraunhofer-Institut für Windenergie und Energiesystemtechnik (IWES)", *SCOPE. Sektorübergreifende Einsatz-und Ausbauoptimierung für Ana-lysen des zukünftigen Energieversorgungssystems*, 2016.
5. "Fraunhofer-Institut für Windenergie und Energiesystemtechnik (IWES)" in Fraunhofer-Institut für Bauphysik (IBP) Institut für Energie- und Umweltforschung (ifeu) and Stiftung Umweltenergierecht Interaktion Eli-Strom. Wärme und Verkehr. Analyse der Interaktion zwischen den Sektoren Strom WärmelKiilte und Verkehr in Deutschland im Hin-blick auf steigende Anteile fluktuierender Erneuerbarer Energien im Strombereich unter Berücksichtigung der europäischen Entwicklung Ableitung von optimalen Entwicklungspfaden für den Verkehrs-und Wärmesektor, Kassel:Heidelberg, Würzburg, 2015.
6. A. von Oehsen, *Entwicklung und Anwendung einer Kraftwerks-und Speichereinsatzoptimierung für die Untersuchung von Energieversorgungsszenarien mit hohem Anteil erneuerbarer Energien in Deutschland*, 2012.
7. M. Jentsch, *Potenziale von Power-to-Gas Energiespeichern*, 2014, ISBN 978-3-8396-0865-4.

8. P. Hartel and M. Korpas, "Aggregation Methods for Modelling Hydropower and Its Implications for a Highly Decarbonised Energy System in Europe", *Energies*, vol. 10, pp. 1841, 2017.
9. T. Trost, *Erneuerbare Mobilität im motorisierten Individualverkehr*, 2016, ISBN 978-3-8396-1129-6.
10. T. Trost, M. Sterner and T. Bruckner, "Impact of electric vehicles and synthetic gaseous fuels on final energy consumption and carbon dioxide emissions in Germany based on long-term vehicle fleet modelling", *Energy*, vol. 141, pp. 1215-1225, 2017.
11. M. Pfennig, N. Gerhardt, C. Pape and D. Böttger, *Mittel-und langfristige Potenziale von PtL-und H2-Importen aus internationalen EE-Vorzugsregionen*, 2017.
12. N. Gerhardt, M. Jentsch, M. von Bonin, S. Becker and D. Böttger, "Entwicklung des Straβenverkehrs und Rückkopplung mit dem Energiesystem in −95% THG-Klimazielszenarien", *Teilbericht im Rahmen des Projek-tes: KLIMAWIRKSAMKEIT ELEKTROMOBILITAT - Entwicklungsoptionen des Straßenverkehrs unter Berücksichiigung der Ritckkopplung des Energieversorgungssystems in Hinblick auf mittel- und langfristige Klimaziele*, 2018.

Chapter - 26

Design of Electric Power Spot Market Mechanism to Encourage the Use of Renewable Energy

Authors

Avik Datta
Electrical Engineering, Swami Vivekananda University,
Kolkata, West Bengal, India

Rituparna Mitra
Electrical Engineering, Swami Vivekananda University,
Kolkata, West Bengal, India

Rituparna Mukherjee
Electrical Engineering, Swami Vivekananda University,
Kolkata, West Bengal, India

Promit Kumar Saha
Electrical Engineering, Swami Vivekananda University,
Kolkata, West Bengal, India

Titas Kumar Nag
Electrical Engineering, Swami Vivekananda University,
Kolkata, West Bengal, India

Susmita Dhar Mukherjee
Electrical Engineering, Swami Vivekananda University,
Kolkata, West Bengal, India

Chapter - 26

Design of Electric Power Spot Market Mechanism to Encourage the Use of Renewable Energy

Avik Datta, Rituparna Mitra, Rituparna Mukherjee, Promit Kumar Saha, Titas Kumar Nag and Susmita Dhar Mukherjee

Abstract

The spot market is essential for advancing the development of the energy market and enhancing its security and effectiveness. To achieve geographic representation, energy savings, and environmental preservation, a low-consumption and high-efficiency power market is suggested. To take into account the key characteristics of the surrounding area and address the key issues with the energy market, the construction of the spot market is explained. The framework for the spot market is recommended to encourage the consumption of renewable energy for the power grid with a high proportion of renewable energy based on the experience of the domestic and international power markets.

Keywords: Spot market, renewable energy consumption, high proportion of renewable energy power grid.

Introduction

China pledged to reach the peak of carbon emissions and consume 20% non-fossil energy by 2030 in the context of the third energy revolution [1]. The National Development and Reform Commission and the National Energy Administration outlined the goals and conditions for the development of the electric power spot market in 2017 and chose eight regions, including the province of Gansu, to carry out a pilot project [2]. Countries in Europe and America already have a lot of experience building electric power spot markets.

The US PJM uses the "full power optimization" model to encourage competition and takes into consideration the actual physical model to guarantee the system's safe and dependable operation while taking into

account both the market's effectiveness and the power grid's security [3].

Long-term physical trading is prevalent on the Nordic regional multilateral spot market. The TSOs of various nations ensure the equilibrium of power supply and demand, successfully coordinate the various resource characteristics of various nations, and eventually transform the market from a decentralized to a centralized one [4]. The expansion of the local power market has merged with the spot market building plans announced by several Chinese provinces to maximize the benefits of regional spot market construction. The southern spot market building plan, which begins in Guangdong, calls for the gradual transition from a pilot spot market to a regionally united spot market based on long-term financial contracts [5].

Zhejiang Province plans to establish a provincial power market system with the spot market serving as the mainstay and the finance market serving as a supplement [6]. This is due to the province's developed market economy, rapid development of direct trading on both sides of power generation, and wealth of experience operating in the power market. The growth of the spot market should choose appropriate construction routes, design plans in accordance with the stage of regional development, and establish fair construction targets based on the source and network structure, supply and demand condition, and degree of marketization. This study analyzes the key features of power development in high-energy regions based on the development of domestic and foreign spot markets, creates the spot market mechanism to encourage the consumption of renewable energy, and validates the market.

Principal Component Analysis (PCA)

a) Introduction

In recent years, China's wind turbines and photovoltaics have experienced spectacular growth thanks to the progressive strengthening of the electricity system and the encouragement of regulations. The structure of the power supply and the condition of supply and demand in these regions have changed due to the rapid development of renewable energy power generation, creating new issues for the market development of regions with a high concentration of new energy sources. The installed energy capacity is expanding quickly in regions with a large supply of renewable energy sources, outpacing the rise in local electricity consumption by a wide margin. As a result, there is a plentiful supply of electricity and there is a

persistent surplus. According to China's economic and energy characteristics, regions with low levels of economic growth have abundant energy.

b) Problem statement

The template is used to format your paper and style the text. All margins, column widths, line spaces, and text fonts are prescribed; please do not alter them. You may note peculiarities. For example, the head margin in this template measures proportionately more than is customary. This measurement and others are deliberate, using specifications that anticipate your paper as one part of the entire proceedings, and not as an independent document. Please do not revise any of the current designations. In light of the fundamental traits of places with a high proportion of renewable energy, the growth of the power market mostly faces the same issues.

First, the significant trend of abandoning wind and solar is the worst issue. The main factors limiting energy consumption are lack of room for energy consumption, inadequate power grid peaking capacity, and difficulty in power grid control. The usage of renewable energy has been somewhat constrained by the ongoing decline in the electrical market. The inconsistent output of renewable energy sources and the absence of peak marketization processes have increased the power of inadequate grid peaking capacity.

Additionally, weak AC grids and a significant amount of new energy have made power grid control more challenging. For instance, the Northwest Power Grid gave up 4.714 billion kWh of solar power and 16.871 billion kWh of wind power in 2018. Both the wind rejection rate (16.21%) and the light rejection rate (8.91%) were significantly higher than the national averages of 7.2% for the wind rejection rate and 3.0% for the light rejection rate.

The coordination strategy for renewable energy has resulted in significant losses for thermal power firms, and thermal power needs to be enhanced.

The major body of thermal power has changed from being the primary source of electricity to being the primary source of capacity supply in recent years due to the cleanliness of the power supply structure. However, as the price of carbon slowly rises, the cost of producing thermal power rises, and the market for thermal power becomes more constrained as a result of the forced abolition of renewable energy sources. The interests of thermal power and renewable energy cannot be appropriately balanced by the market's current solutions since they are not reliable and perfect enough.

Clustering method and indicator

a) Introduction

It is advised to construct a decentralized power spot market with partial power competition in light of the significant issues high-energy areas face and the state of the market. Spot transactions take place in the day-ahead, intraday, and balance markets whereas long-term trades involve physical contracts.

Electric energy, power generating, and auxiliary services are the major aims of the transaction [7, 8].

b) Electrical energy trading market

Renewable energy power generation companies can organize electric energy spot transactions in three stages: day-ahead market, intraday market, and balancing market after applying the long-term decomposition plan transactions. In the day-ahead market, the electricity sales firms and wholesale consumers announce the electricity demand price curve on the working day, and the renewable energy power producing enterprises submit the daily electricity price quote information. The day-ahead market employed centralized bidding and marginal clearing, with each 15-minute period being considered a clearing stage for transactions. The power dispatching agency develops the boundary conditions of the transaction based on the external power supply in the form of an intergovernmental framework agreement and a national power distribution plan, which takes into account the annual contract, interprovincial market-oriented transaction results, renewable energy consumption requirements, and power grid safety operation requirements. Social welfare maximization is the optimization goal while taking into account load forecasting, out-of-sale electricity sales curve, generator operation constraints, power grid safety operation constraints, and other factors. The safety constraint unit combination (SCUC) and safety constraint economic scheduling (SCED) algorithms are used for optimization calculation. The time-sharing power generating output curve, the unit startup combination, and the daily electricity pricing should all be clear on the operating day. The estimate for renewable power has a significant degree of error in the day-ahead market due to the enormous installed capacity of renewable energy and the participation of more and more renewable energy in the market. The intraday trading gives the market a platform to modify the daily trading strategy, which is mostly utilized to address the issue of

significant variance. Intraday trading is structured around a predetermined time frame. The intraday market is launched after the day-ahead market ends and runs for 2-4 hours every 15 minutes for the following hour until the real-time operation. The rolling correction revision process was implemented in the daytime electric energy trading to give the new energy power generation enterprises priority in adjusting the electricity price quotation plan (the new energy enterprises can apply for a downward adjustment after a more accurate power forecast, but it is not possible to apply for an increase in market output). The trading plan is then updated regularly, adjusted, and announced in accordance with the revised quotation plan of the renewable energy generation firm, the updated electricity load prediction, and the modified grid operation status. The balancing electric energy market schedules balancing transactions for every five minutes over the following five to ten minutes, primarily addressing the issue of real-time deviation adjustment. The power dispatching agency will clear out the generator set 15 minutes before to the system's real operation based on the daytime market generator set declaration information, which the power sales firm and wholesale users are not required to disclose. The output of renewable energy will be upgraded first and the thermal power will be reduced first in balancing trade, which will result in an increase in the consumption of renewable energy. In 15 minutes, the market for balancing is closed. The twofold settlement mechanism is the foundation of both the markets for auxiliary services and electric energy [9]. The results of the day-ahead market clearing are used to settle the current plan, and the real-time price is utilized to settle the difference between the balancing market and the day-ahead market clearance.

c) Power generation trading

Power generation trading between renewable energy, conventional energy, and coal-fired self-supplied power plants is a market-oriented trading tool that encourages energy conservation and emission reduction, and can play a significant role in actual work [10]. It also helps to achieve energy-saving emission reduction and promote supply-side reform. The expenses of power generating for the entire society can be decreased, social efficiency can be increased, and the environmental goals of renewable energy consumption can be met by substituting power generation in the day-ahead market. The transition of alternative power production from power plant firms with high energy consumption to power plant enterprises with low energy consumption has a specific direction [11]. The power generating rights

of conventional energy power generation companies and self-supplied power plants will be replaced by renewable energy power generation companies. The alternate approach is to swap out sub-cleaning equipment for cleaner ones, and cheap units for expensive ones, or renewable energy power generation for thermal power plants. The thermal power unit is the power generation transferor in the market transaction, whereas the new energy unit is the power generation transferee. When renewable energy sources are able to provide excess electricity, they will replace power generating rights based on the clearing of the electric energy day-ahead market. A two-way declaration (trading time, power amount, price, etc.) is required for the power generation substitution transaction, which uses a centralized quote and matching transaction between the two parties [12]. In its capacity as the party that generates transfer power, the thermal power unit announces the transfer price and the transfer power; in its capacity as the party that generates transferee power, new energy declares the transferee price and the transferee amount. The trading unit that sells for the most money is matched first by the trading unit that sells for the least money [13]. The quote is the average of the sum of the quotations of the transferor and the transferee, and so on, to obtain the clear conclusion, when the transferor's quotation price is larger than or equal to the transferee's quotation. The power generation of the final transaction substitutes the transaction pair, the trading time period, the transaction power, and the price, it is calculated when the safety check of the power dispatching agency is confirmed. The power dispatching agency plans and executes the two parties' alternative transaction, then reports it for filing to the power regulating agency.

d) Peaking auxiliary service market

In recent years, the demands on the system's safe and stable operation have increased as a result of the rapid development of renewable energy, the ongoing growth in the size of power grids, and the significant increase in the complexity of system operation and management [14][15]. Additionally, high-energy zones have substantially greater capability for power supply [16]. Coal-fired power stations are steadily reducing their operating hours. The issues of system peak shaving, heat meeting, and the abandonment of renewable energy are significant [17]. It is challenging for the current auxiliary service structure to match actual demand. Propose to conduct auxiliary service transactions in the spot market in order to further enhance and deepen the compensation mechanism for auxiliary services and to advance the marketization of power auxiliary services. Given that these

regions have substantial installed thermal and new energy capacity, they mostly send power to other power grids. As a result, the deep peak peaking of centralized bidding will be constructed in the first stage [18]. The cost of ancillary services is decided upon in the service market using market-based techniques. The enthusiasm of power producing businesses, power consumers, and power-selling businesses is spurred to take part in auxiliary services, encouraging higher consumption of renewable energy. When a thermal power plant's peaking rate exceeds 50%, an additional service known as "deep peak shaving" is offered to help with peak shaving. The "stepped" compensation and allocation mechanism is used in accordance with the various thermal power unit peaking depths, and the settlement price is determined by the actual clearing price for each file.

Conclusion

This study discusses the main issues with the abandonment of renewable energy and the harm to the interests of thermal power units in the high proportion of renewable energy areas based on the investigation of the establishment of spot markets at home and abroad. Additionally, it suggests the creation of a spot market mechanism with a market structure that takes into account the properties of renewable energy, the creation of market varieties with multiple transaction targets, and other priority renewable energy consumption mechanisms. The promotion effect of using renewable energy is confirmed using the IEEE-118 node model's simulated operation under the planned spot market mechanism. The study presented in this paper is intended to serve as a guide for developing the spot market in areas with a high concentration of renewable energy sources.

References

1. [1] People's Republic of China, United States of America. Sino-US Joint Statement on Climate Change [Z]. 2014-11-12.
2. [2] General Office of the National Development and Reform Commission, Comprehensive Department of the National Energy Administration. Notice on the Pilot Work on the Construction of the Power Spot Market (Development and Reform Office Energy [2017] No. 1453) [EB/OL]. (2017-08-28) [2018-10-27]. http://www.ndrc.gov.cn/gzdt/201709/t20170905_860117.html.
3. [3] P. Zou, QX. Chen, and Q. Xia, *et al*. Logic Analysis of Foreign Power Spot Market Construction and Its Enlightenment and Suggestions to China [J]. Automation of Electric Power Systems, 2014,38(13):18-27.

4. [4] Nordic Energy Regulators. About NordREG 2018[EB/OL]. [2018-10-28]. http://www.nordicenergyregulators.org.
5. [5] China Southern Power Grid Co., Ltd. Work plan for the construction of power spot market in the southern region (Southern Power Grid System [2017] No. 18) [Z]. 2017-08-15.
6. [6] Zhe jiang Provincial People's Government. Zhejiang Province Electric Power System Reform Comprehensive Pilot Program (Zhe Zhengfa [2017] No. 39) [Z]. 2017-09-30.
7. [7] W. Liu, JC.Tan, Current Status and Prospects of Large Users Participating in Power Market Transactions in Southern Regions [J]. China Southern Power Grid Technology, 2017,11(11):68-74.
8. [8] Y. Feng, BP. Zhang, and HW. Liu, *et al*. The demand for electricity in charging facilities participates in the direct transaction mode of large users [J]. China Electric Power, 2018, 51(04): 175-179.
9. [9] C. Dong, W. Huang. American PJM Power Market and Its Enlightenment to Guangdong Power Reform [J]. Yunnan Electric Power Technology, 2017(01): 16-20.
10. [10] X. Zhang, J. Yan, and B. Pang, *et al*. Application and Analysis of Power Generation Trading in Energy Saving and Emission Reduction in China [J]. Automation of Electric Power Systems, 2014, 38(17): 87- 90.
11. [11] CY. Lin, FS. Wen, and WC. Meng, *et al*. Judging Model of Bilateral Trading of Power Generation Rights under Two-Party Price Mechanism [J]. Electric Power Construction, 2016, 37(3): 91-99.
12. [12] Q. Yu. Research on Optimization Model of Regional Power Generation Transaction [D]. Beijing:North China Electric Power University, 2011.
13. [13] SHANG Jin-cheng. Theory and Application of Power Generation Trading Based on Energy Saving and Emission Reduction (I) Theory of Power Generation Trading[J]. Automation of Electric Power Systems, 2009,33(12):46-52
14. [14] YQ. Liu, HP. Zhang, Q. Li, *et al*. Design and Practice of Power Peaking Auxiliary Service Market in Northeast Power Grid[J]. Automation of Electric Power Systems, 2017,41(10):148-154.
15. [15] N. Zhang, TR. Zhou, and CG. Duan, et al. Impact of large-scale wind

farm connecting with power grid on peak load regulation damand [J]. Power System Technology,2010,34(1):152-158.

16. [16] Q. Lü, W. Wang, and S. Han,et al.A new evalution method for wind power curtailment based on system regulation capability[J].Power System Technology,2013,29(7):1887-1894. [17] MS. He Cogeneration unit power peaking operation feasibility study [D]. Beijing: North China Electric Power University, 2012.
17. [18] Q. Zhang, SX. Xin, and JH. Bai, et al. Study of Ancillary Service Market Mechanism for the Promotion of Wind Power Consumption in Norstwest Regin [J]. Electric Power, 2013,46(7):111-115.

Chapter - 27

Analyzing Business Cases for the Integration of the Electric CAR Market Using Cost-Utility Analysis

Authors

Avik Datta
Electrical Engineering, Swami Vivekananda University, Kolkata, West Bengal, India

Rituparna Mitra
Electrical Engineering, Swami Vivekananda University, Kolkata, West Bengal, India

Titas Kumar Nag
Electrical Engineering, Swami Vivekananda University, Kolkata, West Bengal, India

Susmita Dhar Mukherjee
Electrical Engineering, Swami Vivekananda University, Kolkata, West Bengal, India

Promit Kumar Saha
Electrical Engineering, Swami Vivekananda University, Kolkata, West Bengal, India

Suvraujjal Dutta
Electrical Engineering, Swami Vivekananda University, Kolkata, West Bengal, India

Chapter - 27

Analyzing Business Cases for the Integration of the Electric CAR Market Using Cost-Utility Analysis

Avik Datta Rituparna Mitra, Titas Kumar Nag, Susmita Dhar Mukherjee, Promit Kumar Saha and Suvraujjal Dutta

Abstract

Electric vehicle (EV) adoption on a national scale calls for innovative methods of market integration. The controlled charging of EVs provides a trading opportunity, particularly on the reserve power market. A change in the market's role model results from the entry of new players to conduct the business case of trading secondary reserve power. However, not all actors gain from this industry in the same manner. There are some who are better suited. The modified role model, which covers all new market participants involved with EVs, is described in this paper. Additionally, a novel method for determining if it is feasible for various parties to carry out the aforementioned business case is offered. By measuring soft characteristics, the cost-utility analysis determines unique values for each actor and reflects his capacity. This paper analyses the abilities of three different market actors - a car manufacturer, an energy provider, and a grid operator - using a case study.

Keywords: Business case, cost-utility-analysis, electric vehicle, energy market market integration, market players, role model.

1. Introduction

One of the key concepts in both German and European energy strategy is the energy turnaround. The German government is relying on the implementation of smart grid and smart market solutions, as well as the introduction of electric cars (EVs), to meet this goal. Market integration is therefore a key procedure. EVs can be launched using either one of two methods: controlled or uncontrolled charging. The option of controlled charging provides the chance to produce and market goods. Producing market goods by managing EVs means that multiple new participants, such

as the automaker, the operator of the EV pool, and the aggregator, will enter the energy market. The business case for the market integration must be advantageous to all market participants. To describe the viability, a business case analysis is required. A cost-utility analysis methodology was created as a result to evaluate the technical and legal viability of energy economical business cases from the perspective of each market participant. The cost-utility analysis is used to compare several options that are challenging to characterize using quantitative measures. The cost-utility analysis model (CUAM)'s structure, dimensions, and calculation method are all described in the paper. Calculating utilities (individual values) for various players in the German energy market involves using a case study in which EVs are used to produce goods for the reserve market.

The Federal Ministry for the Environment, Nature Conservation, and Nuclear Safety of Germany has provided funding for the research and development initiative known as Controlled Charging V3.0. The goal of the project is to identify a technical and financial optimum for charging EVs inside an intelligent energy management system while taking into account the needs of EV customers, the erratic availability of renewable energies, and the needs of the infrastructure. Making electric car storage capacities available, achieving highly regulated charging operations to maintain overall grid stability, and balancing the erratic availability of renewable energy from sources like wind and PV are key undertakings.

A pool of electric vehicles' ability to supply balancing power on the reserve market is specifically taken into account. A group of seven partners manages the project, which is overseen by Bayerische Motoren Werke AG [1].

2. EVS in the Energy Market Role Model

Role models facilitate the understanding of duties, interactions, connections, and agreements between various parties (such as businesses, systems, or individuals) [2]. The efficiency of all energy-related operations of the energy market (electricity supply and grid stability) is guaranteed with clearly defined duties.

It is crucial to make clear that roles and market actors are different. Different market actors can play different roles, and market actors can play more than one role.

The primary function of the EV is to get the user from point A to point

B. An example of a market actor is a private individual or business that owns the EV and utilises it for transportation needs. The EV market in Germany is still quite new. By the end of 2013, just 6.051 EV had been sold [3]. However, the market will expand, and since there are many EVs, it is advantageous to incorporate them into the energy market. However, the traditional energy role model needs to be modified because it no longer includes EV-related roles. EVs can only supply grid services like reserve power if these new roles have established relationships with and interactions with the others.

a) The current role model

The power generation, trade, retail, transport, and distribution components make up the traditional energy market's value chain. Generator, Retailer, Distribution System Operator (DSO), Transmission System Operator (TSO), Balance Responsible Party, also known as Energy Trader (BRP), Balancing Coordinator, Metering Service Operator, and Grid User are the existing roles in the current role model that perfectly fit in those segments. Generators run conventional and renewable power facilities and sell energy to BRP like energy dealers or industry over the counter or through other marketplaces, such as EEX.

Other services, such as reserve power, can also be sold to the TSO. Retailers have an agreement for the delivery of power with grid users, often known as energy consumers. Additionally, the Retailer has agreements with the DSO that allow them to use the distributional system to supply energy to its customers. The distribution system operator (DSO) is in charge of overseeing and managing the distribution network, which provides energy to grid users including residences and smaller industries with currents under 110kV. By balancing generation and consumption in real time (frequency management), the TSO controls the transmission network and protects the entire power system. Because of this, the TSO purchases balancing power through the reserve power market (regelleistung.net), where potential suppliers of reserve power can advertise their available capacity. Primary, Secondary, and Minute Reserve are three main forms of reserve power that differ in time requirements [4]. The BRP buys and sells energy and has to make sure that the precise equivalence of the bought and sold energy amounts at intervals of 15 minutes results in a balanced balancing group. Is it true that through the Balancing Coordinator, the BRP is required to pay for the balancing energy (reserve power) supplied by the TSO. All balancing groups are coordinated by the balancing coordinator, who also generates

invoices to BRP for the energy consumed during balancing. The installation, use, and upkeep of measurement equipment (metres), as well as metre reading and transfer to DSO and Retailer, are all the responsibility of the Metering Service Operator. Market players play a variety of roles, as was already mentioned. For their respective control zones, the four German TSOs (50 Hz Transmission, Amprion, TransnetBW, and Tennet TSO) simultaneously perform the duties of TSO and Balancing Coordinator.

Additionally, the top four energy suppliers RWE, AEON, Vattenfall, and EnBW perform the functions of Generator, BRP, and Retailer. They generate energy by running their own power plants, purchase energy directly from other power plants, trade energy through marketplaces, and sell and provide energy to grid users utilising a variety of retail companies.

b) The adapted role model

By 2020, the national platform for electromobility projects that there will be 1 million EVs on German roads [5]. These EVs could provide useful services that aid with the grid integration of an increasing amount of renewable energy. The forecasting of feed-in power has inaccuracies because solar-dependent renewable energy sources like photovoltaic have errors. The TSO must close the power deficit, as was already mentioned. When there is a surplus of power, EVs can increase power demand (start charging, increase charging power), and when there is a shortage of power, they can either decrease power demand (stop charging, decrease charging power), or they can supply power to the grid (known as vehicle to grid). The following new roles are necessary to realise such a business: EV, EV User, EV-Service-Provider (EVSP), EV Service Equipment (EVSE or charging station), EV Service Equipment Operator (EVSE Operator or charging station operator), EV-Pool Operator, and an Aggregator (see Figure 1). They will be applied to the business case of reserve power provision via an EV-Pool in order to demonstrate the duties and interactions of these roles.

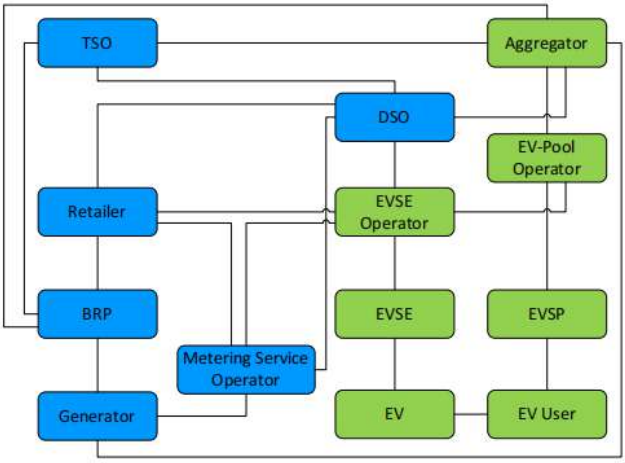


Figure 1: Adopted model

c) The business case reserve power

EVSP provides their clients with EV-related services, such as leasing an EV and a charging station, access to public charging, flexible mobility, and car sharing. Future contracts between EVSP and EV Users on the provision of reserve power are likely. If the provided incentives (such as affordable energy for charging) are compelling enough, EV users will supply their EVs with backup power when they do not need them for transportation (typically at night). When charging at home, EV owners plug their vehicles into their EVSEs and tell the system of the required range and the date and time of their next journey (for instance, through smart phone). The Smartphone sends this information to the EV-Pool Operator, who then receives additional state of charge (SOC) data from the EV.

This data is transmitted directly from the EV through mobile connection to the EV-Pool Operator or via Powerline Communication from the EV through the EVSE and EVSE-Operator Backend. The EV-Pool Operator can determine at any time how much power the EV-Pool can offer to an Aggregator for a specific period of time using this information. The Aggregator hires EV-Pool Operators, grid users (such as cold storage facilities and industrial plants), and generators (such as biomass power plants) and introduces them as an aggregated virtual power plant into the reserve power market. As a result of the combined capacity and the ability to make up for each other's shortcomings, this allows smaller, decentralized consumers and producers to participate in the market. The TSO notifies the

Aggregator if reserve power is needed (negative reserve power). The EV-Pool Operator receives a signal from the Aggregator and requests only those EVs that are able to charge and adhere to the DSO's power restrictions.

Smart meters, which allow for real-time power monitoring over time, will be required for two key goals:

1. To check the requested reserve power
2. To distinguish the energy coming from the retailer or the TSO (negative reserve power) by clearing the Aggregator and Retailer balance groups.

Given that EV users typically shop at different retailers, the clearance between the Aggregator and each Retailer must be completed. Therefore, it makes the procedure simpler if EV customers have access to all the same merchant.

An illustration is provided to highlight the differences between the roles and actors. A full-service provider might give prospective clients with all of their needs from a single source, including contracts for leasing EVs and EVSEs, energy, and reserve power.

The entire service provider also grants the TSO regulation authority at the same time. The actor plays the roles of Retailer, EVSP, EV, EVSE, EVSE Operator, EV-Pool Operator, BRP, and Aggregator in order to realize that business case. Cost-utility analysis is used to assess these types of business propositions.

3. The cost-utility analysis model

The cost-utility analysis model (CUAM) is typically used to assess various investment options that are represented by a range of qualitative criteria [6].

It is difficult to measure these delicate factors. The CUAM is modified for the energy market, and the model is used to evaluate business proposals from the perspectives of various market participants. The CUAM's output expresses each market participant's capacity to carry out the business case. The process and the use of the created CUAM are described in the paragraphs that follow.

a) The methodology

A four-step process is used to calculate each utility while creating a CUAM [6, 7].

- Giving a target system a name. This is a compilation of all criterion required to assess the business case.
- Weighting of the factors. Each market participant assesses each criterion in light of how crucial it is to the implementation of the business case.
- Establishing goals. Each market participant contributes a value to represent how well their business satisfies each requirement.
- Utility cost calculation. The utility for each parameter is derived using weighting factors and goal values. The total utility for the business cases is given by adding together the utilities of all the parameters.

To achieve a full target system that comprises a huge number of parameters to characterise the business cases, a two-sided method is used. Both a top down and bottom up strategy are used to determine the parameters (see Fig. 2).

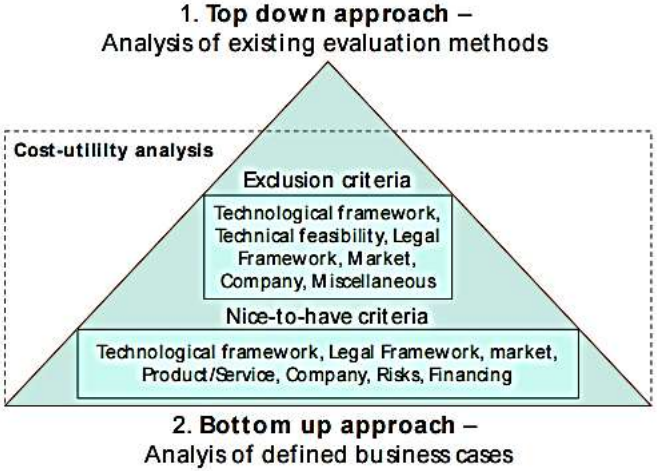


Figure 2: Approach to identify target system

The examination of current target industries for connected business sectors is part of the top-down methodology. The goal is to come up with a general list that makes sense. The issue of assessing business cases in the energy sector is moved to this set of criteria. The list of common characteristics is completed and parameters for specific business cases are identified using the bottom-up methodology [8, 9]. The next paragraph explains how the described strategy was included into the finished CUAM.

b) The implementation

To compute utilities, the CUAM is separated into three sections.

Questions in the first section help create a shared knowledge of the regarded business case. The market participant who wishes to complete the business case provides answers to the questions. The second section identifies the exclusion criteria that, if not met, prevent implementation. The market players' individual assessments will determine the criteria's choice and weighting. There are six categories in the second part. Sub parameters are separated into each group (see table 1).

Table 1: Exemplary list of exclusion criteria

Category	Exclusion criteria
Technological framework	Charging infrastructure, metering device, ITC, controlling devices
Technical feasibility	Interfaces, available data
Legal framework	Laws, directives, contracts, approvals
Market	Market development, market acceptance
Company	Personnel, know-how
Miscellaneous criteria	Available energy source, generation plants

The market participant has the option of selecting non-exclusion criteria or exclusion criteria. A parameter's degree of fulfilment must be entered if it is used as an exclusion criterion. There are two options: yes or no. The market participant has the authority to determine whether or not the criterion is met.

Table 2: Exemplary list of nice-to-have criteria

Category	Nice-to-have criteria
Technological framework	Grid access, locality, Transmission capacities
Legal framework	Legal targets, licenses, approvals
Market	Customer resources, market development, marketing channels, competitive products
Product/Service	Market acceptance, benefit, price
Company	Management qualities, track record, know-how
Risks	Synergistic effects, dependencies, entry barrier
Financing	Internal/external financial resources, accounting methods

4. Case study

The integration of EVs into the energy market necessitates both products that can be sold and responsible market operators. Three actors who can operate electric vehicles to produce goods are chosen for this case study. The reserve power market is where these goods are sold. The next paragraphs contain in-depth information and analysis.

a) General framework

The case study's researched business case is centred on the provision of secondary reserve electricity (see IIC). The CUAM's statements relate to the legislative and technological landscape of the German energy market. To assess the business case, three market participants were picked.

- Car manufacturer
- Energy supplier
- Distribution grid operator (DSO)

Each of the three parties has the ability to exert influence over a specific group of EVs in order to produce goods that meet the demands of the reserve power market. Each market participant serves as an EV-Pool Operator, acting as the intermediary between EV Users, who are the producers, and

Transmission Grid Operators, who are the consumers. Controlled charging is the charging technique in use. The EV-Pool Operator reimburses the EV Users for the capacity their EVs provided.

b) Cost-utility analysis

Market participants can assess their suitability to implement the business case using the CUAM. Table 3 displays the utilities for each category and overall to allow for comparison of the results.

Table 3: Result cost-ultity analysis

Category	Car manufacturer	Energy supplier	DSO
Technological framework	1	0.8	1
Legal framework	0.25	0	0.2
Market	0.52	0.51	0.72
Product/Service	0.61	0.62	0.64
Company	0.8	0.68	0.84
Risks	0.86	0.61	0.78
Financing	0.50	0.59	0.59
Overall	**0.65**	**0.54**	**0.68**

According to Table 2, the DSO with a utility of 0.68 is most suited to carry out the business case. The producer of automobiles comes in second, with a score of 0.65, while the provider of energy comes in third, with a score of 0.54. The limitations of incorporating EVs can be seen when examining the various categories in more detail.

The technology infrastructure is not problematic. The technology requirements can be handled by all actors with ease. In contrast, the major challenges to this business case are related to the legal structure. As of right moment, there are no laws defining a framework for EV market integration. This is highlighted by a poor utility brought on by a persistently low level of fulfilment.

The DSO is most appropriate because of the evaluation of the category "Market." The CUAM claims that the DSO has stronger marketing channels, more customer resources, and higher market volumes. The category

"Products/Service" has an equal evaluation from all three actors, receiving a medium-high score of 0.6. This indicates that the product, known as "controlled charging," is more popular than typical. The distinction between the medium utility of the Energy supplier and the high utilities of the Car manufacturer and the DSO is brought about by the categories "Company" and "Risks." These areas have ratings of 0.8 or higher from both actors. Their businesses are set up for success in the market and can reduce risks.

The medium rating for "financing" indicates that there are ample financial resources to carry out the business case.

The CUAM's findings reflect each market participant's unique viewpoint on how to integrate electric vehicles (EVs) into the energy market by using controlled charging. Legal framework, market, and financing are the three key pillars for a successful integration of EVs. This outcome highlights the necessity of changing the law, opening up the market, and making smart investments.

5. Conclusion

The goal of integrating EVs into the energy market necessitates the development of new market actors and business cases. The modified role model introduces new actors and places them in the existing contractual arrangements. The new roles, in particular the EV-Pool Operator and the Aggregator, are in charge of putting the required business justifications for EVs into action.

Therefore, it makes no difference which actors play the part. A new cost-utility analysis model was created to assess the viability of business cases related to the energy market. The model is a three-step procedure that includes developing the business case, creating exclusion criteria, and grading pertinent business case-related factors. Offering solutions for the reserve power market by using controlled charging for EVs is the business case that is taken into consideration.

This business case will be examined from the perspectives of three market actors. The conclusion drawn from the findings is that the DSO is best positioned to carry out the business case. To better EV integration into the German energy industry, additional changes to the legal system, market access, and finance procedures are needed. The actors in the energy market are only partially represented by the results. As a result, this case study cannot be used to draw general conclusions.

References

1. BMW AG, "Project Flyer Gesteuertes Laden 3.0", Munich, 2013, http://www.erneuerbar-mobil.de/projekte/foerderung-von-vorhaben-imbereich-der-elektromobilitaet-ab-2012/kopplung-der-elektro-mobilitaet-an-erneuerbare-energien-und-deren-netzintegration/projektflyernetzintergration/flyer-gl-3.0.pdf (Mar 2014).
2. CEN-CENELEC-ETSI Smart Charging Working Group, "Report Working Group Smart Charging", 2013.
3. KBA, "Pressemitteilung Nr. 02/2014 - Fahrzeugzulassungen im Dezember 2013 - Jahresbilanz" http://www.kba.de/cln_031/nn_1389648/DE/Presse/Pressemitteilungen/2014/Fahrzeugzulassungen/pm02__2014__n__12__13__pm__text.html (Mar 2014).
4. Amprion GmbH, "Primärregelung, Sekundärregelung, Minutenreserve", http://www.amprion.net/primaerregelung-sekundaerregelungminutenreserve (Mar 2014).
5. Nationale Plattform Elektromobilität (NPE), "Zweiter Bericht Nationale Plattform Elektromobilität", Berlin, 2011.
6. H.J. Wareneck *et al*., "Wirtschaftlichkeitsrechnung für Ingenieure", München, 1996, ISBN 3-446-18593-3.
7. IHK Munich, "Aufbau und Inhalt eines Businessplans", 2013, http://www.muenchen.ihk.de/de/starthilfe/Anhaenge/BusinessplanIHK.pdf (Mar 2014).
8. A. Schneider, "Geschäftsmodellwandel durch disruptive Innovationen Fallstudie zum Elektrofahrzeug in Automobilindustrie und Energiewirtschaft", Driesen, 2012, ISBN 978-3868661316.
9. R. Müller, " Geschäftsmodelle für dezentrale stationäre und mobile Energiespeicher im Bereich Endanwendungen", Master thesis, Ilmenau University of Technology, department Energy Systems, 2011.

Chapter - 28

A Secure Day-Ahead Bidding inside Energy and Reserve Market for a Pulp and Paper Mill

Authors

Avik Datta
Electrical Engineering, Swami Vivekananda University,
Kolkata, West Bengal, India

Rituparna Mitra
Electrical Engineering, Swami Vivekananda University,
Kolkata, West Bengal, India

Rituparna Mukherjee
Electrical Engineering, Swami Vivekananda University,
Kolkata, West Bengal, India

Promit Kumar Saha
Electrical Engineering, Swami Vivekananda University,
Kolkata, West Bengal, India

Titas Kumar Nag
Electrical Engineering, Swami Vivekananda University,
Kolkata, West Bengal, India

Suvraujjal Dutta
Electrical Engineering, Swami Vivekananda University,
Kolkata, West Bengal, India

Chapter - 28

A Secure Day-Ahead Bidding inside Energy and Reserve Market for a Pulp and Paper Mill

Avik Datta, Rituparna Mitra, Rituparna Mukherjee, Promit Kumar Saha, Titas Kumar Nag and Suvraujjal Dutta

Abstract

There is a greater need for reserve capacity because of rising use of vacillating renewable energy foundations. By taking advantage of the built-in flexibility in a few particular industrial processes, non-generating resources like major industrial customers can merchandise energy prices and supply reserve volume. Through efficient bidding in electricity markets, a sizable industrial customer can benefit from this flexibility. The day forward cost abatement of a hazard loath pulp and paper mill (PPM) is presented in this study as a two-stage stochastic issue, taking into account the PPM's thermodynamic and electrical restrictions. The bids are exposed to price unpredictability in the energy and deputy bazaars. The findings of a case study conducted in Sweden show that utilizing PPM's flexibility has a considerable economic impact. Investigated is the hazard hardship of the PPM as well as the estimated rate of the pulp and paper mill consequential since various tactics. We demonstrate how reserve offers can dramatically raise the PPM's profitability.

Keywords: Demand side management, electricity markets, frequency containment reserve, industrial demand response.

1. Introduction

Demand Response (DR) is one of the propitious solutions to the present problems facing the electrical arrangement, rendering to the Nordic TSOs [1]. Sweden can rely on roughly 4,500 MW of variable claim, with the industrial sector accounting for about 51% of it, per [2]. The industries that use a lot of energy are anticipated to have a bigger impact. They are not only the ones who could contribute most effectively, but they are also the ones who would profit most economically from changes in the price of electricity [3]. The

Swedish Energy Agency states that variations in the sector's usage of energy have a substantial influence on the whole industrial energy usage because the Pulp and Paper sectors in Sweden account for around 52% of the total energy consumption [4]. [5] provides a thorough analysis of DR submissions in the manufacturing division. The survey then provides many industries with a larger potential for DR programs after outlining the significance of auxiliary amenities and their possible in various trades. The P&P sector is regarded as having significant DR potential. A pulp and paper mill cogeneration scheme, which usages vapor for 2 resolves to offer heat for the mashing procedure and to produce power for auction to local benefactors presents an energy optimization algorithm by Marshman *et al*. [6]. A dispatch and investment model is created by Santos *et al*. [7] to address the manufacture arrangement and development issues at a combined pulp and paper mill (PPM). In order to improve upon existing models, Figueira *et al*.'s [8] optimization-based choice sustenance approach for a PPM takes into account realistic limitations and goals. Rodriguez-Garcia *et al*.'s simulation implementation [9] for manufacturing users demonstrates the budding of paper manufacture in DR programs as well as the advantages connected to it. The tool is used to do a cost-benefit investigation for the application of Demand Response (DR) stratagems. Lawrence *et al*. [10] examined the steady appearances of the Swedish pulp and paper sector and discovered that, for all types of mills, fluctuation in power costs did not suggestively affect energy competence through the course of the study. The majority of the studies [6-10] has been on increasing the PPM's internal functionality and reducing overall electricity usage, or energy efficiency. The capability of power customers near adjusts their petition to the requirements of the system, including moving demand to other times, is described as demand response (DR) in [5]. The DR probable of a PPM regarding optimized contribution and request in extensive power sooqs has only been examined in a few instances [11, 12]. The broadest foundation of suppleness in the pulp manufacture development, notably in the refiners, was highlighted by those who examined the possibilities of DR in the P&P business. The insignificant volume and high close of utilization of the machines make them the furthermost hopeful ones for harmonizing amenities, according to Paulus *et al*. [11]. The authors also identify a source that can guarantee peak saving on the day-ahead souk in the obtainability of massive pulp stowage. An optimization technique was created by Helin *et al*. [12] to simulate the contribution of certain volume after the refiner unit to the modifiable power souq. This demonstrates that the pulp production stage is thought to have the biggest potential. However,

nobody of the aforementioned publications discusses how a stoppage in the pulp manufacture procedure can lead to a decline in pulp quality. Because of this, we left the refiners out of the study in [13]. Poorer paper is invariably the product of pulp that is of lesser grade. Here, however, we concentrate on the steam accumulator, which adds a large degree of litheness to the invention course by storing steam containing up to 4 MWh of thermal energy.

2. Methodology

a) Pulp & paper production process and model

Spruce logs are debarked and chopped into chips to start the process. When necessary, the substantial is moved from stowage turrets to the tissue sector. Chemical pulping or thermo-mechanical pulping are two methods for producing pulp.

Refiners: In this instance, refiners are making thermomechanical pulp. A stream of timber chips enters refiners to begin the operation. The refiners feature 2 counterrotating corrugated metallic disks that channel the chips toward the boundaries of the disks whereas mining the wood's grit (due to intense centrifugal force). Since significant amounts of energy are lost through friction, this process balance sheet aimed at around 70% of the mill's entire energy ingesting. In the procedure of low-slung pressure haze, about 60% of the power used by the refineries is converted into thermal energy. The water that is present privileged the wood chips vanishes under short pressure as a result of the high temperature inside the refiners. The energy from this "dirty" vapor is then utilised in a heat-exchanger to create unsoiled steam, which is subsequently cast-off in the paper apparatuses.

Bleaching: To improve the pulp's quality in accordance with each client's specific needs, the pulp goes concluded a succession of cleaning and bleaching stages. It does not take a lot of electricity to perform the chemical process of bleaching.

Electric Boilers (B): Two sizable electric boilers can be swiftly triggered to make up aimed at the damage of blistering steam during maintenance or a problem with uncertainty of the steam supplied by the refiners is insufficient aimed at the aeriation process. The electric ratings of the two separate boilers are 60 MW and 40 MW, respectively. The steam generated through the boilers can subsequently be used in the paper machines, held in the accumulator, or spilled.

Steam Accumulator (A): PPM under investigation, a vapor accumulator is extant. It can store pressured saturated liquid at pressures of upto 13 bar

with a thermal energy plasticity of 4 MWh (heat content equal to 6-10 MWh of thermal energy).

Auxiliary 10 bar grid (C): A steady stream of steam is drawn from the steam accumulator and hassled to 10 bar, where it is castoff for supplementary operations.

Spillage (S): Steam might spill from the low-pressure bar if it has produced too much of it.

Paper Machines (PM): The paper machines employ steam that has been "cleaned" by refiners, added mist as of electric boilers, and prospective haze that may be unconfined by the haze accumulator. A dewatering step is the first in the wet end. Wet pulp is sprayed onto a small, revolving net by a large, flat nozzle, which then moves the substantial to the unrelenting area. Here, extra aquatic is mislaid and the final product's breadth is changed. The residual water in the paper is then completely removed by the now "clean" steam as the sheet moves through the drying stage. The paper is prepared to remain wound onto spools and then cut to fulfill a specific shopper order. Understanding Flexibility, the electric boiler and steam accumulator were the loads with the greatest degree of flexibility. The energy magnitude of the haze accumulator besides the possibility of steam spills supports the power measurements of the electrical boilers, which can lone be utilized. The electric boilers are incredibly quick to adapt power usage in both directions (up- and downwards) and may ramp to full power in a matter of seconds. The (60 + 40) MW electric rating indicates a sizable amount of potential power capacity. In the event that neither the refiners nor the boilers are able to supply the necessary steam, the steam accumulator was set up as a backup to prevent a steam shortage. It can also be utilized to store 4 MWh of thermal energy.

b) Nordic electricity and reserve markets

This study looks at a PPM that engages in diurnal ahead energy and fallback market as well as implicitly settling imbalances because it is a BRP. The subsequent markets, they are identified by their superiors as below, are available for bids by the PPM.

DA: The day-ahead energy spot market where energy e_t^{DA} is purchased at price λ_t^{DA} at hour t.

RT: Imbalance settlement for real-time energy consumption. A one-price consumption imbalance settlement [14] is assumed where the PPM pays λ_t^{RT} for deviation of the real-time consumption e_t^{RT} from the DA purchase.

R : Frequency containment reserve in normal operation (FCR-N), i.e. primary frequency regulation, where capacity p_t^R is remunerated with λ_t^R.

I : Imbalance fee λ^I that penalizes uninstructed energy deviations from the DA bids in real-time.

The total cost of the PPM at hour t is thus

$$e_t^{DA} \cdot \lambda_t^{DA} + (e_t^{RT} - e_t^{DA}) \cdot (\lambda_t^{RT} + \lambda_t^I) - p_t^R \cdot \lambda_t^R \quad \text{SEK.} \quad (1)$$

c) Assumptions

The condensed plant diagram is shown in Fig. 1. Steam after various pressure echelons is changed to the thermal energy q in period interlude t, which is displayed in bloodshot, while mass movements are emphasized in lime. Electric authority and energy are specified in blue. The following set of presumptions makes it easier to derive the condensed PPM model [13] shown in Fig. 1.

- The thermal energy that steam has in its mass stream per hour is used to represent it. We calculate thermal energy present in hourly vapor flow q using the temperature, pressure, and heat content of wet steam while disregarding losses associated with steam expansion (isenthalpic procedure).
- The electrical boilers can be instantly controlled.
- The following day's regular production taxes are known.
- The daily revenue from the sale of papers is consistent.
- The PPM accurately predicts the prices of reserves and energy.
- A price taker model of the PPM is possible.
- No minimum bid amount is required. The marginal system price is acceptable for energy and reserve bids.
- The real-time initiation of standby energy is adequately represented with hourly mean values of frequency, and stand-in bids are paid for power capacity regardless of the genuine real-time energy instigation.

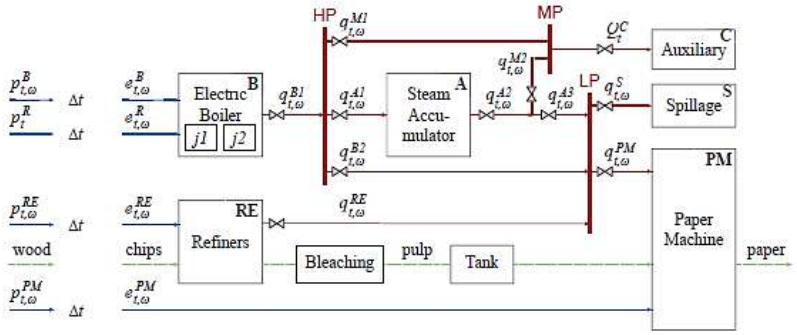


Figure 1: Simplified plant scheme of a pulp & paper mill (from [13]) in Sweden. Electric power *p* and energy *e* are indicated in blue, steam from different pressure levels is converted to the thermal energy *q* in time interval Δ*t* in red, and mass flows are highlited in green

c) Problem formulation

As a 2-phase stochastic issue in [13] with price uncertainty for RT prices and frequency realizations, we define the cost minimization of the risk-averse PPM. The goal of the PPM is to reduce the provisional value at risk (CVaR), which is the weighted average of predicted cost.

$$\begin{aligned} \min. \quad & (1-\beta)\cdot \mathbb{E}\left[c_\omega^T\right] + \beta \cdot \mathrm{CVaR} \qquad (2) \\ \text{s.t.} \quad & \mathbb{E}\left[c_\omega\right] = c^{DA} - c^R + \sum_{\omega\in\Omega} \pi_\omega \left(c_\omega^{RT} + c_\omega^I + c_\omega^X\right), \end{aligned}$$

Cost and revenues, Operational penalties,
Electric power & energy constraints, Steam balance,
Start-up & shut-down of el. boilers, Production limits,
Accumulator balance, Line limits, Risk-aversion

Table 1: Definition of the five summer and winter cases

Case	Description
1 Reference	w/o Accumulator, w/o Reserve Provision
2 Energy Optimization	w/ Accumulator, w/o Reserve Provision
3 Energy & Reserve	w/ Accumulator, w/ Reserve Provision
4 No Accumulator	w/o Accumulator, w/ Reserve Provision
5 No Spillage	high spillage penalty 5000 $\frac{\text{SEK}}{\text{MWh}}$

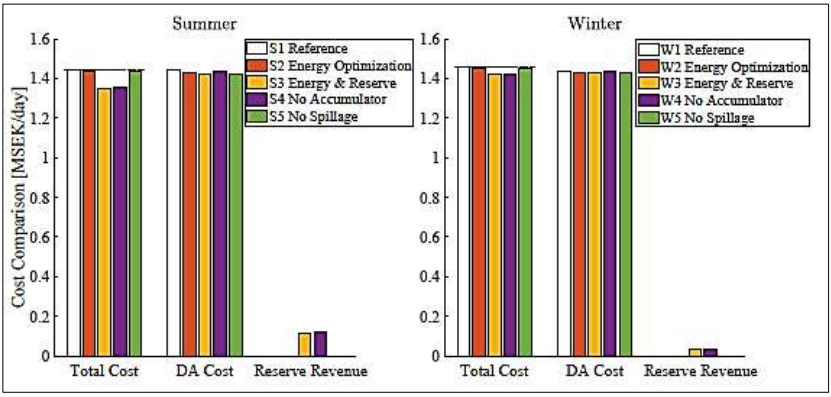


Figure 2: Cost comparison of case studies in terms of total expected cost, day ahead cost for energy purchase, and revenue from reserve capacity

3. Results

In Sweden's price region SE3, the archetypal is functional to a peril opposed (= 0.6) PPM with ancient fee and occurrence set-ups. Public statistics are available for the characteristics relating to the energy market; hourly energy charge information is provided by Nord Pool [15] besides reserve amounts by Mimer [16]. You may get the old frequency facts at Fingrid [17]. The statistics pertaining to PPMs was gathered after a PPM in Sweden. We use 300 scenarios in the summer and 300 situations in the winter to prototypical real-time inequity defrayal values and incidence understandings. Table I provides a summary of the case study's input parameters. As the reference scenario, scenario 1 represents the status quo in which a fixed amount of electricity is purchased, no optization is carried out, and the accumulator is not used for plasticity. Energy arbitrage is described in Case 2 after the PPM simply improves their DA get-up-and-go bids and real period consumption with the goal of utilizing variations in energy prices during different hours. A PPM that concurrently enhances energy and replacement offers is described in Case 3. Case 4 simulates a PPM without a steam accumulator that concurrently optimizes energy and reserve bids. Case 5 simulates a PPM that works to prevent steam spills while jointly optimizing energy and reserve bids. Fig. 2 displays the final total cost, DA cost, and reserve income. Cost savings are possible in every situation and are often greatest in the summer. This can largely be related to increased summertime reserve prices and associated earnings. In the summer, 0.4%, 6.5%, and 6.0% can be saved in cases 2 and 5, respectively. Figure 3

displays the outcomes of the summer cases S2 through S5. Figure 3a shows the outcomes in the energy arbitrage instance. According to the scenario, the "violin plots show" the probability density of second-stage decisions, and the dots indicate the expected value. The top left graph demonstrates that there is no backup power available and that the electric boiler's power is heavily varied based on the outcome of the RT pricing. The market's bids for energy are displayed below. "While the RT" consumption is at least 195.3 MWh/h, the DA bid swings between 192 and 203 MWh/h. This meets the PPM's minimum energy and steam requirements along "with 4 MWh" from "the steam accumulator." Depending on the situation, the steam accumulator is frequently used and is predicted to be at 50% at the end of the day. The electric boilers use more energy in the rare cases where overall consumption surpasses 204 MW, and extra steam is released. This is only cost-effective "in a one-price imbalance" settlement system "when the system" is "down regulating" and demands an increase in demand. Throughout the day, the anticipated steam spillage is rather modest "and only" increases in situations where the boiler uses too much energy. "The RT" energy usage in "the boiler is" displayed on the graph in the top right corner. "The daily" RT consumption, "reserve capacity," spillage, and "stored steam" are "compared to" the reference case in the bottom right bar graph. Figure 3b shows the "outcomes of joint" energy and reserve optimization. Throughout the day, "a total of" 406 MW of "reserve power capacity" is made available, although the predicted value of the activation energy for reserves is almost zero. Even more spillage results from "the additional" reserve bids, "their stochastic" and the boiler's excessive" energy use.

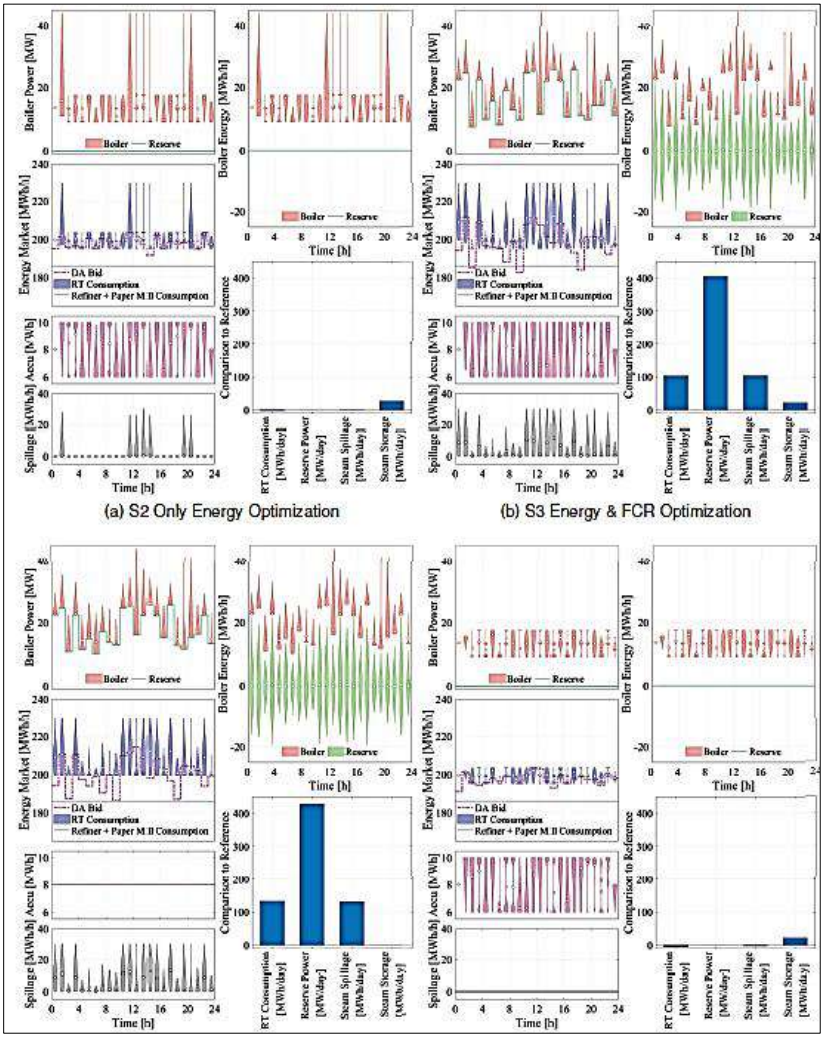


Figure 3: Result of selected summer cases. Violin plots show the probability density of second-stage decision while white dots show the expected value

The outcomes "of joint" energy and "reserve optimization" without the accumulator are shown in Fig. 3c. The reserve provision is comparable in size, if not somewhat greater.

The outcomes are displayed in Fig. 3d when spillage is prohibited but the accumulator is used. A more cautious DA energy bidding technique is used, and no reserves are offered.

In this instance, the accumulator's flexibility is not employed to supply reserves, but rather to arbitrage energy prices.

4. Discussion

The main source of reserve capacity flexibility “in this” case study is steam spillage. “The flexibility” is provided by the quick electric boilers, “and the” spillage allows for a wide “range of” power “consumption in” the “electric boiler” without affecting the PPM steam grid. This finding initially appears counterintuitive given that wasting energy runs counter to all energy saving measures. However, power capacity, not energy, is the primary issue with the integration of renewable energy.

5. Conclusion

In this study, we used “a simplified” model of a “Pulp-&-Paper” Mill (PPM) to determine the best day-ahead bids in the reserve and power markets while taking risk aversion into account. We show how the suggested simplified PPM model may be used for day-ahead bidding with uncertainty. Prices for “imbalance settlement” and “real-time frequency” realization (as “hourly averages”) are both predicated on uncertainty. The operational cost savings are anticipated to be in the range of 6.5%, or 93,000 SEK “per day.” We discover “that the” steam accumulator plays a little part in cost reductions while offering “flexibility for” energy arbitrage. It is not necessary to supply reserve” capacity, however. Show, however, “that the” fast-acting electric boiler and steam spills are a significant source of flexibility for reserve capacity and cost reduction.

References

1. “Challenges and Opportunities for the Nordic Power System,” Statnett, Tech. Rep., 2016.
2. L. S¨oder, P. D. Lund, H. Koduvere, T. F. Bolkesjø, G. H. Rossebø, E. Rosenlund-Soysal, K. Skytte, J. Katz, and D. Blumberga, “A review of demand side flexibility potential in Northern Europe,” *Renewable and Sustainable Energy Reviews*, vol. 91, no. March, pp. 654-664, 2018.
3. “Measures to increase demand side flexibility in the Swedish electricity system,” Swedish Energy Markets Inspectorate, Tech. Rep., 2017. [Online]. Available: https://www.ei.se/PageFiles/308322/Ei*{ }*R2017*{}*10.pdf
4. Swedish Energy Agency, “Energy in Sweden (2017),” Swedish Energy Agency, Tech. Rep., 2018. [Online]. Available: https: //energimyndigheten.a-w2m.se/

5. M. H. Shoreh, P. Siano, M. Shafie-khah, V. Loia, and J. P. Catal˜ao, "A survey of industrial applications of Demand Response," *Electric Power Systems Research*, vol. 141, pp. 31-49, 2016.

6. D. Marshman, T. Chmelyk, M. Sidhu, R. Gopaluni, and G. Dumont, "Energy optimization in a pulp and paper mill cogeneration facility," *Applied Energy*, vol. 87, no. 11, pp. 3514-3525, 2010.

7. M. O. Santos and B. Almada-Lobo, "Integrated pulp and paper mill planning and scheduling," *Computers & Industrial Engineering*, vol. 63, no. 1, pp. 1-12, 2012.

8. G. Figueira, P. Amorim, L. Guimar˜aes, M. Amorim-Lopes, F. Neves-Moreira, and B. Almada-Lobo, "A decision support system for the operational production planning and scheduling of an integrated pulp and paper mill," *Computers & Chemical Engineering*, vol. 77, pp. 85-104, 2015.

9. J. Rodr´ıguez-Garc´ıa, C. A´ lvarez-Bel, J.-F. Carbonell-Carretero, M. Alc´azar-Ortega, and E. Pe˜nalvo-L´opez, "A novel tool for the evaluation and assessment of demand response activities in the industrial sector," *Energy*, vol. 113, pp. 1136-1146, 2016.

10. A. Lawrence, M. Karlsson, and P. Thollander, "Effects of firm characteristics and energy management for improving energy efficiency in the pulp and paper industry," *Energy*, vol. 153, pp. 825-835, 2018.

11. M. Paulus and F. Borggrefe, "The potential of demand-side management in energy-intensive industries for electricity markets in Germany," *Applied Energy*, vol. 88, no. 2, pp. 432-441, 2011.

12. K. Helin, A. K¨aki, B. Zakeri, R. Lahdelma, and S. Syri, "Economic potential of industrial demand side management in pulp and paper industry," *Energy*, vol. 141, pp. 1681-1694, 2017.

13. L. Herre, F. Tomasini, K. Paridari, and L. S¨oder, "A simplified model of a pulp-& paper mill for optimal bidding in electricity wholesale markets and primary frequency reserve provision," *Applied Energy submitted*, 2019.

14. eSett Oy Settlement Service, "eSett Handbook," 2018. [Online]. Available: https://www.esett.com/handbook/

15. Nord Pool AS, "Historic Market Data," October 2018. [Online]. Available: http://www.nordpoolspot.com/historical-market-data/

16. Svenska Kraftn¨at, “Mimer - Primary Regulation Prices.” [Online]. Available:https://mimer.svk.se/PrimaryRegulation/PrimaryRegulationIndex
17. Fingrid, “Fingrid - Open Data.” [Online]. Available: https://data.fingrid.fi/en/dataset/frequency-historical-data

Chapter - 29

Power System and Faults: A Review

Authors

Rituparna Mukherjee
Electrical Engineering, Swami Vivekananda University,
Kolkata, West Bengal, India

Susmita Dhar Mukherjee
Electrical Engineering, Swami Vivekananda University,
Kolkata, West Bengal, India

Abhishek Dhar
Electrical Engineering, Swami Vivekananda University,
Kolkata, West Bengal, India

Promit Kumar Saha
Electrical Engineering, Swami Vivekananda University,
Kolkata, West Bengal, India

Sujoy Bhowmik
Electrical Engineering, Swami Vivekananda University,
Kolkata, West Bengal, India

Arunima Mahapatra
Electrical Engineering, Swami Vivekananda University,
Kolkata, West Bengal, India

Chapter - 29

Power System and Faults: A Review

Rituparna Mukherjee, Susmita Dhar Mukherjee, Abhishek Dhar, Promit Kumar Saha, Sujoy Bhowmik and Arunima Mahapatra

Abstract

In an electric power system, a fault or fault flow is any unusual electric flow. For instance, a short circuit is a fault where current detours the typical load. An open circuit fault happens if a circuit is hindered by some disappointment. In three-stage systems, a fault might include at least one stages and ground, or may happen just between stages. In a "ground fault" or "earth fault", current streams into the earth. The forthcoming short-circuit current of a predictable fault can be determined for most circumstances. In power systems, defensive gadgets can recognize fault conditions and operate circuit breakers and different gadgets to restrict the deficiency of administration because of a disappointment. This paper reviews about the power systems and power faults.

Keywords: Power system, faults.

1. Introduction

The power system is a network of millions of electrical segments working in synchronism. The current, voltage, power and recurrence are the fundamental boundaries of the power system. In ordinary condition, these all boundaries are stays in appraised esteem. In strange condition or faulty condition, these boundaries are cross their constraints of evaluated esteem. The power system partitions into three sections; power age, power transmission, and power dissemination. One all the more part you can add that is insurance and switchgear [1].

The electricity or electric power creates in power plants. The electrical power plants convert some other type of energy into electrical energy. The other wellspring of energy might be compound energy, sun-based energy, wind energy, thermal power. As indicated by the accessibility of sources, various kinds of power plants use in the network. The power plants partition

into two sections; sustainable power plants and non-inexhaustible power plants [2].

The inexhaustible power plants utilize a wellspring of energy like sun based, wind, hydro, biomass and geothermal. The non-sustainable power plants utilize a wellspring of energy like coal, atomic, flammable gas, diesel. The nuclear energy station creates the greater part of the electrical energy. Be that as it may, we need to discover other elective fuel of hotspots for what's to come. These producing stations place at a far separation from the load or city. The transmission network uses to communicate power over a huge distance [2].

2. Power transmission

Power plants place at a far separation from the city region or last customer. The transmission system is a connection between creating system and appropriation system. It has the biggest region contrast with power age and power appropriation system. The transmission network is generally inclined to fault. There are more opportunities to fault happen in the transmission line. The huge limit of the transmission line conveys a lot of loads. In the event that this transmission line will fall flat, an exceptionally enormous measure of load will separate from the system. Along these lines, insurance of a transmission line is generally significant. Three kinds of transmission lines are accessible as indicated by the length of the line; short transmission line, medium transmission line and long transmission line. Under 60 km of the line is consider as a short transmission line. 60 km to 250 km line is consider as a medium transmission line and in excess of 250 km line is consider as long transmission line. The high voltage transmission lines use to lessen copper misfortune and size of conductor. For the most part, the voltage level of the AC transmission line is 230 kV, 500 kV, 765 kV and for DC transmission lines and 100 kV to 1500 kV [2].

3. Power distribution

The voltage level of the circulation network is under 132 kV. The transmission line of 11 kV, 33 kV, 66 kV, and 132 kV consider in the circulation system. These lines convey less power contrast with the lines of transmission networks. The power system network supplies electrical power to the private, business and modern load. It supplies 3 stage power for modern load and the voltage level is 440 V. It supplies single stage power for business and private load and voltage level is 230 V [2].

Requirement of electrical power system according to a financial perspective, we generally build a producing station where assets are promptly accessible. Customers devour electrical energy; however, they might remain in such locations where the assets for delivering electricity are not accessible. Not just that, occasionally there are numerous different limitations because of which we can't develop a creating station closer to the thick shopper's territories or load focuses [3].

So, all things being equal we utilize a remotely found age source and afterward send this produced power to the load habitats through a long transmission line and a dissemination system. We call the whole course of action from producing plants to buyer closes for conveying electricity productively and dependably as the electric power system [3].

4. Faults in electrical power system

Electrical networks, machines and supplies are frequently exposed to different kinds of faults while they are in activity. At the point when a fault happens, the trademark esteems (like impedance) of the machines might change from existing qualities to various qualities till the fault is cleared.

There might be part of probabilities of faults to show up in the power system network, including lighting, wind, tree falling on lines, device disappointment, and so forth.

Figure 1: Electrical faults

A fault in an electric power system can be characterized as, any unusual state of the system that includes the electrical disappointment of the hardware, for example, transformers, generators, busbars, and so on.

The fault initiation likewise includes in protection disappointments and directing way disappointments which results short circuit and open circuit of conductors.

Under typical or safe working conditions, the electric supplies in a power system network operate at ordinary voltage and flow appraisals. When the fault happens in a circuit or gadget, voltage and current qualities goes astray from their ostensible reaches. The faults in power system causes over current, under voltage, unbalance of the stages, turned around power and high voltage floods. This outcomes in the interference of the typical activity of the network, disappointment of supplies, electrical fires, and so forth Normally power system networks are ensured with switchgear insurance types of gear, for example, circuit breakers and relays to restrict the deficiency of administration because of the electrical disappointments. The fault in the power system is characterized as the imperfection in the power system because of which the current is occupied from the planned way. The fault makes the unusual condition which lessens the protection strength between the conductors. The decrease in protection makes over the top harm the system. The fault in the power system is principally sorted into two kinds they are

1. Open Circuit Fault
2. Short Circuit Fault.

The various sorts of power system fault are displayed beneath in the picture:

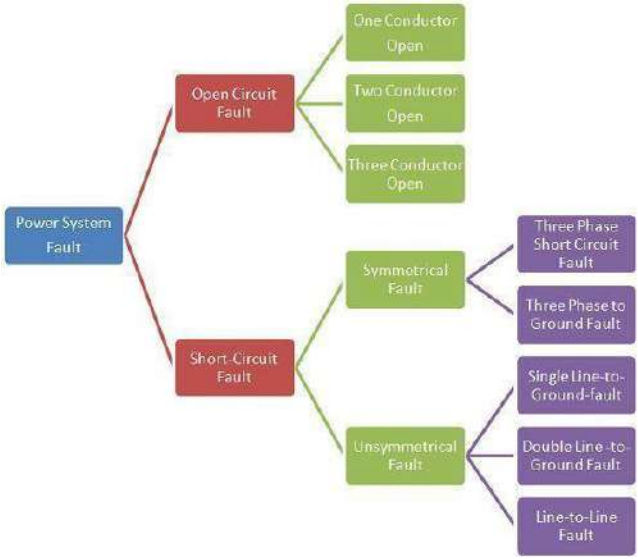


Figure 2: Fault types

The faults in the power system might happen in view of the quantity of regular unsettling influences like lightning, rapid breezes, earthquake, and so forth It might likewise happen as a result of certain mishaps like tumbling off a tree, vehicle crashing, with supporting design, plane slamming, and so forth.

4.1 Open circuit fault

The open circuit fault primarily happens on account of the disappointment of a couple of conductors. The open circuit fault happens in series with the line, and along these lines, it is likewise called the series fault. Such kinds of faults influence the unwavering quality of the system. The open circuit fault is sorted as

- Open conductor fault
- Two conductors open fault
- Three conductors open fault.

The open circuit fault is displayed in the figure beneath.

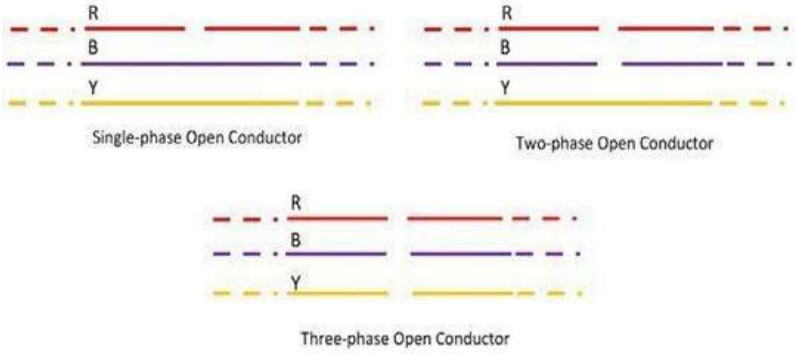


Figure 3: Open circuit faults

4.2 Short-circuit fault

In this kind of fault, the conductors of the various stages come into contact with one another with a power line, power transformer or some other circuit component because of which the huge current stream in a couple of periods of the system. The short-circuit fault is separated into the balanced and unsymmetrical fault.

Even fault

The faults which include every one of the three stages is known as the

even fault. Such kinds of fault stay adjusted even after the fault. The even faults principally happen at the terminal of the generators. The fault on the system might emerge by virtue of the opposition of the curve between the conductors or because of the lower balance obstruction. The even fault is sub-classified into line-to-line-to-line fault and three-stage line-to-ground-fault

a) Line - Line - Line Fault - Such sorts of faults are adjusted, i.e., the system stays even after the fault. The L - L - L fault happens seldom; however, it is the most serious kind of fault which includes the biggest current. This huge current is utilized for deciding the rating of the circuit breaker.

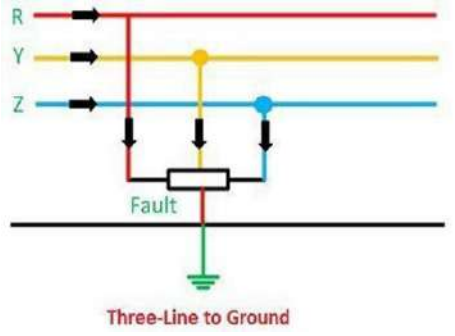


Figure 4: Line-line-line symmetrical fault

b) L - L - L - G (Three-phase line to the ground fault) -The three-stage line to ground fault incorporates all the three period of the system. The L - L - L - G fault happens between the three stages and the ground of the system. The likelihood of event of such kind of fault is almost 2 to 3 percent.

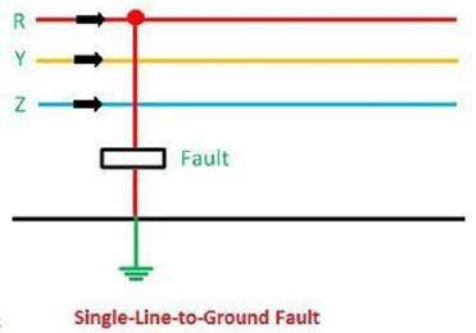


Figure 5: Single-line-to-ground symmetrical fault

Conclusion

In power systems, defensive gadgets can recognize fault conditions and operate circuit breakers and different gadgets to restrict the deficiency of administration because of a disappointment. This paper reviews about the power systems and power faults.

References

1. J. Ma X. Yan B. Fan C. Liu and J. S. Thorp "A novel line protection scheme for a single phase-toground fault based on voltage phase comparison" IEEE Trans. Power Del. vol. 31 no. 5 pp. 2018-2027 Oct. 2016.
2. E. C. Piesciorovsky and N. N. Schulz "Fuse relay adaptive overcurrent protection scheme for microgrid with distributed generators" IET Gen. Transm. Distrib. vol. 11 no. 2 pp. 540-549 Jan. 2017.
3. F. B. Costa A. Monti and S. C. Paiva "Overcurrent protection in distribution systems with distributed generation based on the real-time boundary wavelet transform" IEEE Trans. Power Del. vol. 32 no. 1 pp. 462-473 Feb. 2017.
4. V. C. Nikolaidis E. Papanikolaou and A. S. Safigianni "A communi-cation-assisted overcurrent protection scheme for radial distribution systems with distributed generation" IEEE Trans. Smart Grid vol. 7 no. 1 pp. 114-123 Jan. 2016.
5. F. Xu, W. Huang, L. Zhou, N. Tai, J. Wen and L. Cao, "An intermittent high-impedance fault identification method based on transient power direction detection and intermittency detection," 2017 IEEE Power & Energy Society General Meeting, 2017, pp. 1-5.
6. P. Xi, P. Feilai, L. Yongchao, L. Zhiping and L. Long, "Fault Detection Algorithm for Power Distribution Network Based on Sparse SelfEncoding Neural Network," 2017 International Conference on Smart Grid and Electrical Automation (ICSGEA), 2017, pp. 9-12
7. J. Kumar and P. Jena, "Fault detection in a series compensated line during power swing using superimposed apparent power," 2015 IEEE Power & Energy Society General Meeting, 2015, pp. 1-6.
8. T. K. Patel, S. K. Mohanty and S. Mohapatra, "Fault detection during power swing by phase difference technique," 2017 Innovations in Power and Advanced Computing Technologies (i-PACT), 2017, pp. 1-6.

9. T. Wu, G. Tu, Z. Q. Bo and A. Klimek, "Fuzzy Set Theory and Fault Tree Analysis based Method Suitable for Fault Diagnosis of Power Transformer," 2007 International Conference on Intelligent Systems Applications to Power Systems, 2007, pp. 1-5.
10. A. Pradhan and G. Rao, "Differential power based symmetrical fault detection during power swing," 2013 IEEE Power & Energy Society General Meeting, 2013, pp. 1-1.
11. L. Zamboni, I. Nunes da Silva, L. Nascimento Soares and R. A. Souza Fernandes, "Fault Detection in Power Distribution Systems Using Automated Integration of Computational Intelligence Tools," in IEEE Latin America Transactions, vol. 9, no. 4, pp. 522-527, July 2011.

Chapter - 30

A Review on Fault Detection on DC Microgrid

Authors

Rituparna Mukherjee
Electrical Engineering, Swami Vivekananda University,
Kolkata, West Bengal, India

Susmita Dhar Mukherjee
Electrical Engineering, Swami Vivekananda University,
Kolkata, West Bengal, India

Abhishek Dhar
Electrical Engineering, Swami Vivekananda University,
Kolkata, West Bengal, India

Promit Kumar Saha
Electrical Engineering, Swami Vivekananda University,
Kolkata, West Bengal, India

Sujoy Bhowmik
Electrical Engineering, Swami Vivekananda University,
Kolkata, West Bengal, India

Arunima Mahapatra
Electrical Engineering, Swami Vivekananda University,
Kolkata, West Bengal, India

Chapter - 30

A Review on Fault Detection on DC Microgrid

Rituparna Mukherjee, Susmita Dhar Mukherjee, Abhishek Dhar, Promit Kumar Saha, Sujoy Bhowmik and Arunima Mahapatra

Abstract

Due to inherent advantages of DC system over AC system such as compatibility with renewable energy sources, storage devices and modern loads, Direct Current Microgrid (DCMG) has been one of the key research areas from last few years. In the proposed work power has been produced by using solar pv and fuel cell which is combined on the same dc bus and given to a dc load. DC microgrid protection is an important part of power system studies. The basic think which is required for protection is to get to know, where it should be provided and which component must be used. So, to address the challenges of dc microgrid protection accurate fault detection strategies, fault current limiting method and dc circuit breaker has been used. The electricity load of the microgrid is satisfied by the power form fuel cell, pv array and battery altogether. Initially battery is charged up to 70% and at this time bidirectional converter work as a boost converter and when battery discharges till 30%, battery starts charging and then bi-directional converter act as buck converter and battery remain in charging mode.

Keywords: Micro grid, fault detection, fault classification, renewable energy.

Introduction

A microgrid is a local electrical grid containing sources and loads. The micro grid is a cluster of distributed resources which have the ability to operate autonomously. Micro grids are being used to bring electricity into areas where transmission lines cannot reach. A traditional system with generation in one place and then distribution at high voltages is designed for high energy density fossil fuels. Distributed generation can be used to increase the reliability of a system and allow for the integration of renewables. Distributed generation is a much more suitable method of

electricity distribution for renewables due to their lower energy density as compared to fossil fuels and since the power generation is on site losses due to transmitting electricity are proportionally eliminated. Energy storage can be used in micro grids to improve the power quality and smooth out the fluctuations of renewable energy generation. The recent trend in renewables is to use distributed power sources and energy storage to form a micro grid. DC micro grids are not very widespread but have the potential to present many advantages in terms of facilitating renewable energy integration and improving power quality. DC micro grids usually contain distributed energy resources (DER), loads and energy storage. Renewable energy sources such as photovoltaic modules and wind turbines are typically connected to the DC bus via power electronic converters. These converters have the ability to control the output voltage of DER in order to stabilize the bus voltage and extract maximum power. There are power electronic converters that have the ability to increase or decrease the output voltage. DC loads can be directly connected to the DC bus and if an AC load is required an inverter would be needed in order to invert the DC bus voltage into a usable AC voltage. Batteries are typically used in DC micro grids due to their relatively cheap price and longer backup times. A longer backup time and low losses are desirable for energy storage technologies for micro grids which contain renewable energy generation in order for the load to be met. The problem with batteries is that their service life is relatively short and therefore they need to be changed out more often. Charge controllers are used in order to control the flow of power in the micro grid. Devices are needed to control when power is sent to the batteries or sent to power the load. These controllers also help to improve the power quality in the micro grid. Why fault occur in DC micro grid? Due to the low impedance nature of DC micro grid system, the capacitive filters associated with converters will rapidly discharge into a fault, resulting a large current surge within very short duration.

The increased fossil fuel usage has resulted in increase in greenhouse gases and an alarming reduction in the fossil fuel resources. Here comes into picture, the importance of micro grid (MG), as an economically viable, clean and efficient power source. MGs are the low voltage grids, which connect group of micro sources along with the storage devices. These micro sources generate electricity at distribution level. This is called distributed generation (DG) [1]. As the DG penetration increases, the unpredictable behaviour of the micro sources complicates the inter operation of DG with the distribution

system. To overcome this problems, the DG's are connected to the MG with the help of power electronics based converters. This inverter based distributed generators (IBDGs), due to their inherent current control techniques, limit the fault current to 2 p.u. during the islanded mode of operation. This restriction imposes a cascading effect on protection of distribution system by commonly used cost-effective overcurrent relays (OCR), which renders the fault being undetected [2-6]. These varying operating conditions and restriction in magnitude of the fault current make modification of over current relays an imperative requirement for smooth functioning. The MG protection issues like bidirectional power flow, protection blinding, sympathetic tripping, OCR selectivity & sensitivity and limitation of fault current leads to the demand of better and more effective protection schemes.

Solutions lie either in pre-designing the relays, factoring the operating conditions or by real time altering of relay settings as per the emergence of faults and location of DGs. In pre-designing of relays, multiple options exist of any particular operation, which enables the use of different optimization techniques such as adaptive modified firefly algorithm [6], teaching learning-based algorithm [7], symbiotic organism search optimization [8], and particle swarm optimization [9]. In utilizing the second option, that is adaptive real time setting, use is made of an approximate Thevenin equivalent circuit [10]. A linear programming based online coordination algorithm for relay adaptive protection is illustrated in [11]. One of the most important contributing factors towards OCR dependability and security is their coordination. As part of constant learning and getting adapted to varying network configurations, use of ANN is made in [12]. Relay coordination by utilizing multi-agent is given in [13]. OCR coordination by hybrid algorithm for radial network is proposed in [14]. Back up protection necessitates development of a digital protection scheme, while keeping in mind the relay time grading coordination. TMS and PSM settings are optimized in a microgrid model by utilizing particle swarm optimization (PSO) [15]. Machine learning is emerging out as yet another effective tool for adaptive protection of OCRs. This method aims at intelligent modification of OCR settings for varying operating conditions [16]. A hybrid method known as cuckoo linear optimization algorithm has been used for optimal protection coordination of OCRs. This hybrid algorithm converges to optimum value at much faster rate as compared to PSO or GA. Another attempt has been made in [17] for achieving adaptive overcurrent protection, by continuous

monitoring of the micro grid, using information collected from different relays by Micro grid Control Centre.

Another attempt has been made in [21], where an adaptive relay using curve fitting technique has been proposed. This scheme proposes optimal 3D-curves in respect of selected relays by employing MATLAB curve fitting tool. In this paper, the adaptive relay performance is also compared with traditional OCRs and has been found more efficient, as far as protection issues are concerned. A novel time current voltage characteristic has been proposed in [22]. It uses not only current but voltage magnitude as well and has been found to reduce the overall time of operation of DOCR's, for meshed distribution network. In another approach [23], a dual setting DOCR was proposed. Here every DOCR has 2 relay settings depending on fault direction. It was found that the time of operation was reduced by almost 50% as compared to conventional method, irrespective of size and location of DG. A novel technique is proposed in [24], to design and analyses DOCR protection with proper settings, in both the modes of microgrid operation.

a) Differential protection

A number of researchers have aimed at achieving microgrid protection with the aid of differential protection. The basic concept of operating principle of the differential protection relays has been elucidated in [25]. A suggested method of symmetrical components of current based protection scheme is also elucidated, employed for detection of L-G, LLG faults in microgrids [26]. Another protection scheme making use of direction and magnitude of current employed for the protection of a microgrid, comprising both synchronous and inverter based DGs is explained in [27]. For the protection of both radial and meshed microgrids, a scheme based on instantaneous current differential protection is elucidated in [28]. In addition, the method has a capability of high impedance fault (HIF) detection. Another attempt has been made in [29], where a differential scheme based on spectral energy at either end is proposed. The rationale behind tripping remains same as differential schemes based on current, where, relay trips when the energy differential exceeds the defined threshold. However, identification of the threshold poses some challenges as it depends on multiple factors, like fault impedance and operating modes to name a few. A current differential protection scheme utilizing power line carrier communication technique for protection of meshed microgrids is elucidated in [30]. Another current differential scheme with communication channel for islanded/ grid

connected microgrid protection is shown in [31]. The implementation of fuzzy processes (FHSP) with hilbert space-based power theory, has been used for microgrid protection using differential scheme. The underlying concept of functioning of this scheme is the power differential at either end of the feeder. The advantages accrued by this process include fault detection within two cycles from fault incidence as also the quality of detection being unaltered by threshold values and validity for all kinds of faults including internal faults.

Major advantages of differential protection schemes are that functioning of these schemes is unaffected by the bidirectional flow of power in grid connected mode or by lessening of fault current in islanded connection. Some shortcomings of the differential scheme can be listed as high dependence on communication performance, requirement of a dedicated communication link, saturation or current transformer ratio mismatch.

b) Voltage based scheme

These utilise voltage magnitude as an index and were proposed to protect the MG. The three phase voltages are transformed to dq reference frame using Park transformation and then the value of positive sequence voltage is assessed in comparison to that in dq reference frame, to identify the fault in the MG. Both symmetrical and unsymmetrical faults for highly penetrated system are detected by this scheme. In [32], DG output voltages are transformed to dq reference frame, to protect in-zone and out-zone faults in MG. Major disadvantages of voltage-based scheme are its sensitivity towards voltage changes, dependency on MG configuration and limitation for HIF detection.

c) Communication based protection scheme

A microprocessor-based relays incorporating directional elements can be used to protect a MG in both the modes of operation. In [28], digital relay communication network is used based on current differential scheme to protect a MG. It is applicable for HIFs detection. Digital relays are found quite effective for detection and isolation of faults in a MG, but at the same time are very costly due to communication network involved. It also requires some features which are absent in present time protection system.

d) Distance protection

In case of distance protection, admittance or impedance measurement is used as a parameter to detect faults. New types of admittance relay utilizing inverse time tripping characteristics is proposed in [33]. This admittance relay

is able to identify faults in both the operating modes of a MG. In this scheme inverse time characteristics is to be added to each protection zone and at the same time it should be able to detect both forward and reverse faults. The scheme suffers with higher tripping time as a result of downstream source in feed, decreased accuracy due to current transients, reducing DC magnitude and time constant. Distance protection exhibits very less changes in its features for either of the operation mode of the MG as compared to over current protection. Relays show better coordination and effectiveness in the MG network.

e) Risks associated with short circuit currents

The building/facility may not be properly protected against short-circuit currents. These currents can damage or deteriorate equipment. Improperly protected short-circuit currents can injure or kill maintenance personnel. Recently, new initiatives have been taken to require facilities to properly identify these dangerous points within the power distribution of the facility.

Why short circuit is dangerous?

Short circuit current can be very large. If unusually high currents exceed the capability of protective devices (fuses, circuit breakers, etc.) it can result in large, rapid releases of energy in the form of heat, intense magnetic fields, and even potentially as explosions known as an arc blast. The heat can damage or destroy wiring insulation and electrical components. An arc blast produces a shock wave that may carry vaporized or molten metal, and can be fatal to unprotected people who are close by. Therefore, reliable protection scheme should be developed for the DC microgrid acting in both grid and islanded mode of operation. The objective of this paper is to propose a model for short circuit fault detection and protection of DC microgrid consisting of renewable energy generation. A DC microgrid model has been designed and simulated that comprises a protection model for multiple energy sources and equipments. The components of the DC microgrid and the protective devices will be modelled then whole DC microgrid will finally be simulated.

i) DC microgrid protection

In addition to salient protection challenges, in AC micro grids, DC micro grid is inherent with quick discharge of capacitor which manifests in a higher value of fault current. In order to mitigate any damages that can be caused to electronic devices, a very small-time frame of 2ms must be utilized to isolate the micro grid. Current and voltage derivatives are much better

utilized than pure current and voltage-based methods, responses of earlier being more sensitive [34]. Elucidates a current derivative based fault detection method. Although the current and voltage change based fault detection methods are good enough for fault detection, however, these are easily affected by the inherent transient behaviour of power system and the measuring noise. Schemes based on differential protection are much better solutions for protection of DC microgrid [35]. The vector difference between post fault and pre fault currents, termed as superimposed current, can be used as a parameter to design protection scheme. Differential scheme uses a greater number of relays and measuring instrument, which adds to the protection cost. Protection of DC microgrids with ring configuration can be achieved, where an intelligent device is used, to attain increased reliability and decreased protection cost. Moreover, the scheme was found to identify low as well as high impedance faults in few milliseconds. Table 1. Below elucidates comparison between the functioning of various protection schemes in a microgrid.

ii) Conclusion

With the advent of faster communication systems, adaptive protection has emerged as one avenue which gives out a broad spectrum of possibilities towards the development of new age protection systems. This has opened doors for adaptive protection implementation in AC MGs. Differential protection is also a very good solution for protection of AC micro grids, as it is unaffected by bidirectional power flow in a MG. Other avenues towards addressing the protection challenges are implementation of voltage based and distance protection Schemes, which can be explored further. In DC MG, as fault has to be isolated quickly, differential protection becomes an obvious choice. The idea of hybrid AC-DC microgrids, having benefits over pure AC or DC MG is coming up recently. The development of hybrid AC-DC micro grid protection is still in nascent stage and there lies much scope of research in this direction.

Table 1: Comparative analysis of various protection schemes as applied to microgrids

Parameter technique	Bidirectional	MG operation mode	Communication medium	Relay coord	Tripping time
OCR protection	Highly affected	Highly affected due	Required for Adaptive OCR	Gets adversely	Faster

		to reduced currents in islanded mode			
Differential protection	Not affected	Not affected	Highly dependent	Good	Fast
Voltage based protection	Not affected	Depends on MG mode	Not dependent	<<Gets affected	Fast
Communication based protection	Not affected	Not affected	High cost medium	Very good	Fastest
Voltage based protection	Not affected	Depends on MG mode	Not dependent	<<Gets affected	Fast

References

1. S. Mirsaeidi, D. M. Said, Md. W. Mustafa, and M. H. Habibiddin, “Progress and problems in micro-grid protection schemes,” Renew. Sustain. Energy Rev., vol. 37, pp. 834- 839, Sept. 2014.
2. S. Parhizi, H. Lotfi, A. Khodaei and S. Bahramirad, "State of the artin research on microgrids: A Review," IEEE Access, vol. 3, pp. 890925, June 2015.
3. B. J. Brearley and R. R. Prabu, “A review on issues and approaches for microgrid protection,” Renew. Sustain. Energy Rev., vol. 67, pp. 988-997, Jan 2017.
4. L. F. De. F. Gutierres, L. Mariotto, G. Cardoso, and F. Loose "Recloser-fuse coordination protection for inverter-based distributed generation systems" in Proc. 50th International Universities Power Engineering Conference (UPEC), UK, pp. 1-6, Sept. 2015.
5. Y. Paithankar, Fundamentals of Power System Protection, PHI Learning Private Limited, New Delhi, pp.1-50, 2014.
6. A. Tjahjono, D. O. Anggriawan, A. K. Faizin, A. Priyadi, M. Pujiantara T. Taufik, and M. H. Purnomo, "Adaptive modified firefly algorithm for optimal coordination of over current relays," IET Generation. Transmission. Distribution, vol. 11 no. 10, pp. 2575-2585, July 2017.
7. M. Singh, B. K. Panigrahi, and A. R. Abhyankar "Optimal coordination of directional over-current relays using Teaching Learning-Based Optimization (TLBO) algorithm," Int. J. Electr.Power Energy Syst., vol. 50, no. 1, pp. 33-41, Sept. 2013.

8. A. Datta, D. Saha, and P. Das "Optimal coordination of directional overcurrent relays in power systems using Symbiotic Organism Search Optimission technique" IET Generation. Transmission. Distribution, vol. 10, no. 11, pp. 2681-2688, May 2016.
9. H. Zeineldin, E. F. El-Saadany, and M. M. A. Salama, "Optimal coordination of over current relays using a modified particle swarm optimization," Electrical. Power Syst. Res., vol. 76, no. 11, pp. 988-995, Jan. 2006.
10. Fonseca, Antonio & Salazar, G. & Quilumba, Franklin & Perez-Yauli, Fabian. (2020). Determination of the Thevenin equivalent in power grids using real records based on short circuit power. IET Generation, Transmission & Distribution. 15. 10.1049/gtd2.12003.
11. S. Shen, D. Lin, H. Wang, P. Hu, K. Jiang, D. Lin, and B. He, "An adaptive protection scheme for distribution systems with DGs based on optimized Thevenin equivalent parameters estimation," IEEE Trans. Power Delv., vol. 32, no. 1, pp. 411-419, Feb. 2017.
12. B. Chattopahyay, M. S. Sachdev, and T. S. Sidhu, "An online relay coordination algorithm for adaptive protection using linear programming technique," IEEE Trans. Power Del., vol. 11 no. 1 pp. 165-173, Jan. 1996.
13. E. Koley, R. Kumar, and S. Ghosh "Electrical power and energy systems low cost microcontroller based fault detector classifier zone identifier and locator for transmission lines using wavelet transform and artificial neural network: A hardware co-simulation approach" Int. Electrical. Power Energy Syst., vol. 81, pp. 346-360, Oct. 2016.
14. E. Baran and I. M. El-Markabi "A multiagentbased dispatching scheme for distributed generators for voltage support on distribution feeders," IEEE Trans. Power Syst., vol. 22, no. 1, pp. 52-59, Feb. 2007.
15. A. A. Kida, and L. A. Gallego, "A high performance hybrid algorithm to solve the optimal coordination of over current relays in radial distribution networks considering several curve shapes," Electrical. Power Syst. Res., vol. 140, pp. 464-472, 2016.
16. H. S. Pandya, D. M. Pandeji, R. K. Iyer, and P. M. Purohit, "Digital protection strategy of microgrid with relay time grading using particle swarm optimization," 5th Nirma University International Conference Engineering (NUiCONE), Ahmedabad, 2015, pp. 1-6.

17. H. Lin, K. Sun, Z. Tan, C. Liu, J. M. Guerrero, and J. C. Vasquez, "Adaptive protection combined with machine learning for microgrids," IET Generation Transmission Distribution, vol. 13, no. 6, pp. 770779, March. 2019.

18. N. Kumar and D. K. Jain, "An Adaptive Inverse-Time Over current Protection Method for Low Voltage Microgrid," 2020 IEEE 9th Power International Conf (PIICON), 2020,1-5.

19. A. Darabi, M. bagheri, and G. B. Gharehpetian, "Highly sensitive microgrid protection using over current relays with a novel relay characteristic," IET Renew. Power Generation, vol. 14, no. 7, pp. 12011209, March 2020.

20. A. Srivastava, J. M. Tripathi, Abhinav and S. K. Parida, "Protection coordination in grid connected and islanded mode of microgrid," in Proc. IEEE 8th Power India International Conference (PIICON), Kurukshetra, India, 2018, pp. 1-6.

21. K. Gupta and S. Sarangi, "Adaptive over current relay setting for distribution system using superconducting fault current limiters," in Proc. IEEE 8th Power India International Conference (PIICON), Kurukshetra, India, 2018, pp. 1-6.

22. V. Vinod and U. J. Shenoy, "An adaptive relay based on curve fitting technique for micro-grid protection," in Proc. IEEE Int. Conf. on Power Elect, Smart Grid and Renew. Energy. (PESGRE2020), Cochin, India, 2020, pp. 1-6.

23. H. H. Zeineldin, H. M. Sharaf, D. K. Ibrahim and E. E. A. El-Zahab, "Optimal Protection Coordination for Meshed Distribution Systems With DG Using Dual Setting Directional Over-Current Relays," IEEE Trans. Smart Grid, vol. 6, no. 1, pp. 115-123, Jan. 2015.

24. P. Naveen and P. Jena, "A Review on Issues and Coordination Strategies for Over Current Protection in Microgrid," in Proc. 14th IEEE India Council International Conference (INDICON), 2017, pp. 1-6.

25. N. Patil, M. K. Kirar, P. Paliwal and S. K. Wankhede, "Protection of Microgrid Using Coordinated Directional Overcurrent and Undervoltage Relay," 2021 International Conference on Sustainable Energy and Future Electric Transportation (SEFET), 2021,1-5.

26. M. Dewadasa, A. Ghosh, and G. Ledwich, "Protection of microgrids using differential relays," AUPEC 2011, pp. 1-6.

27. H. Nikkhajoei, and R. H. Lassester, "Microgrid fault protection based on symmetrical and differential current components," University of WisconsinMadison, Dec. 2006.

28. S. Conti, L. Raffa, and U. Vagliasindi "Innovative solutions for protection schemes in autonomous MV micro-grids," in Proc. Int. Conference on Clean Electrical Power, 2009, pp. 647-654.

29. E. Sortomme, S. S. Venkata and J. Mitra, "Microgrid Protection Using CommunicationAssisted Digital Relays," IEEE Trans. Power Deliver., vol. 25, no. 4, 2789-2796, Oct. 2010.

30. S. R. Samantaray, G. Joos, and I. Kamwa, "Differential energy based microgrid protection against fault conditions," IEEE PES Innovative Smart Grid Tech. (ISGT),1-7, 2012.

31. L. Mansinha, R. Stockwell, and R. Lowe, "Pattern analysis with two dimensional spectral localisation: Applications of twodimensional S transforms," Physics A: Statistical Mechanics and its Appl., vol. 239 no. 1-3, pp. 286-295, 1997.

32. M. Dewadasa, A. Ghosh, and G. Ledwich, "Protection of microgrids using differential relays," in Proc. Univ. Power Eng. Conf., 2011, pp.1-6.

33. H. Al-Nasseri, and M. A. Redfern, "A new voltage based relay scheme to protect microgrids dominated by embedded generation using solid state converters," in Proc. Int. Conf. on Electricity Distri.,2007, pp. 1-4.

34. G. Buigues, A. Dysko, V. Valverde, I. Zamora, and E. Fernández, “Microgrid Protection: Technical challenges and existing techniques,” International Conference on Renewable Energies and Power Quality (ICREPQ’13), 2013, vol. 1, no. 11, pp. 222227.

35. A. Meghwani, S. C. Srivastava, and S. Chakrabarti, "A new protection scheme for DC microgrid using line current derivative," IEEE Power & Energy Society General Meeting, Denver, CO, 2015, pp. 1-5.

36. R. Mohanty and A.K. Pradhan, "A Superimposed Current Based Unit Protection Scheme for DC Microgrid," IEEE Trans. on Smart Grid, vol. 9, no. 4, pp. 3917-3919, July 2018.

Chapter - 31

A Review on Fault Detection and Allocation Techniques in Power System

Authors

Rituparna Mukherjee
Electrical Engineering, Swami Vivekananda University, Kolkata, West Bengal, India

Susmita Dhar Mukherjee
Electrical Engineering, Swami Vivekananda University, Kolkata, West Bengal, India

Abhishek Dhar
Electrical Engineering, Swami Vivekananda University, Kolkata, West Bengal, India

Promit Kumar Saha
Electrical Engineering, Swami Vivekananda University, Kolkata, West Bengal, India

Suvraujjal Dutta
Electrical Engineering, Swami Vivekananda University, Kolkata, West Bengal, India

Titas Kumar Nag
Electrical Engineering, Swami Vivekananda University, Kolkata, West Bengal, India

Chapter - 31

A Review on Fault Detection and Allocation Techniques in Power System

Rituparna Mukherjee, Susmita Dhar Mukherjee, Abhishek Dhar, Promit Kumar Saha, Suvraujjal Dutta and Titas Kumar Nag

Abstract

Electric faults are a significant challenge in today's large-scale power systems; they can affect the system's reliability, health, and continuity of the power supply. Faults in the power system must be found, assigned, and fixed in order to provide appropriate power supply continuity and dependability. Electrical problems can be caused by both human irresponsibility and natural disasters. Defect identification and allocation using the traditional method required a significant amount of time and work. As a result, numerous academics create diverse fault allocation and detection techniques in this reference. These tactics are crucial to the smart grid's transition from the conventional grid. An overview of several recently created cutting-edge fault detection and fault allocation techniques is given in this work. These techniques can be utilized in power system for provide protection to the power system equipment. This paper helps researchers to find the fault detection and allocation techniques in single place.

Keywords: Power system fault, Smart grid, Fault detection techniques.

1. Introduction

The current electrical grid was created more than a century ago, when electricity demands were restricted [1]. Power generation was limited and concentrated around settlements, and most residences only required minimum amounts of it [2]. Grid size also grows daily in accordance with energy requirements. Hence, there is a greater chance of various failure types in large-scale grids. Both natural and man-made factors may have contributed to these flaws. The following figure 1 lists the various kinds of errors.

Resolving faults or abnormal conditions in the system is necessary to

achieve proper reliability and continuity. The current, conventional methods for identifying, allocating, and fixing faults need more labour and time.

Numerous academics have developed a variety of fault detection and allocation techniques in an effort to address this issue. These methods are essential to the transition from a traditional to a smart grid. This paper provides a brief overview of a few of these methods.

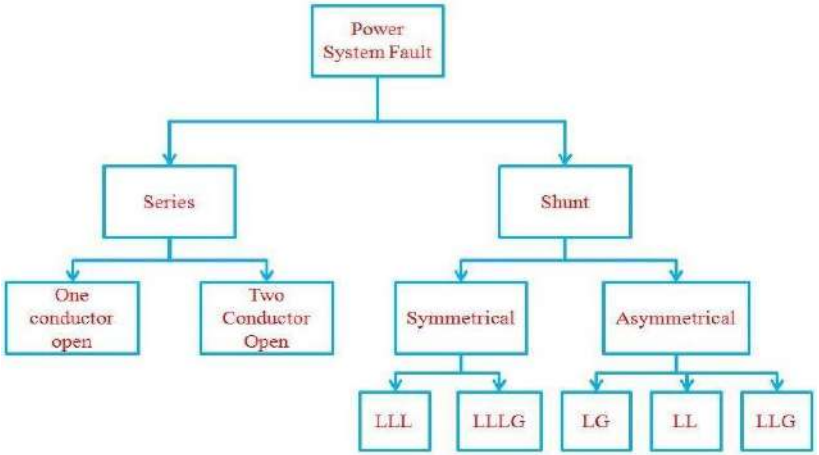


Figure 1: Power system fault

2. Smart grid

An electrical grid featuring a number of advanced metering features, such as integrated demand response and home management through smart distribution boards and circuit breakers, is referred to as a smart grid.

- Load control switches and smart appliances, which are typically funded by efficiency upgrades to local initiatives
- The capacity to recharge larger battery arrays constructed from these or other energy storage devices, as well as batteries from renewable energy sources, when an electric vehicle is parked.
- Energy-conserving assets
- Distribution of electrical surplus through power lines and auto smart switches
- Enough utility-grade fibre broadband, with wireless backup, to link and monitor the aforementioned. Ample reserve of "dark" capacity, frequently rented for profit, to guarantee failover.

The development of electrical regulation and electronic power conditioning and distribution are key elements of the smart grid. While the term "smart grid" usually refers to the technological infrastructure, a fundamental re-engineering of the electricity services industry is also part of the adoption of smart grid technology. The majority of concerns regarding smart grid technology are focused on smart meters, the products they enable, and overall security issues.

3. Fault detection and allocation techniques

The brief description of some novel fault detection and allocation techniques are gives as under:

An accurate fault location (FL) technique based on impedance is: This method's fundamental calculation is the voltage sag between two measuring points, which are selected in accordance with the network topology, the strategic measurement points of the line, and the substation's current readings. The suggested approach is evaluated on a radial power system with various types of faults to show its accuracy and effectiveness [3].

i) Locating the Two-Terminal Fault Combination: The method combines fault location approaches for natural frequencies and travelling waves. In the first step of the walking wave natural frequency approach, the approximate position range of the fault site is calculated to find the estimated time range of the fault's initial travelling wave to reach the bus bar. After that, the second step of accurate ranging is finished by using the Ensemble Empirical Mode Decomposition (EEMD) algorithm to break down the processed signal. From there, the high-frequency travelling wave component is extracted, the initial head recognition and calibration travelling wave is identified, the time parameters required for fault ranging are obtained, and the more accurate fault point location is determined using the two-ended fault ranging formula. We produced a simulation. We created a simulation model of the MMC-HVDC system line fault in PSCAD/EMTDC, and the simulation experiments are compared with different fault types and locations. The results show that the combined fault ranging method effectively identifies the wave head and eliminates the influence of wave speed on accurate positioning, further improving ranging accuracy [4].

ii) LV side measurement-based approach for allocating faults: Based on the distributed measurement at the low-voltage (LV) side, a fault

location technique is suggested. There are clear variations in the voltage variation (phase voltage and sequence voltage) at the LV side for various faults based on the analysis of voltage distribution rules and the distribution transformer's impact on voltage transmission. More precisely, there are variations in the voltage's detection sensitivity to various faults. As a result, a fault location technique based on characteristic voltage at the LV side is suggested for the MV distribution network. A genuine distribution network and a simulation of the IEEE 34 system constructed in PSCAD/EMTDC were used to confirm the accuracy and efficacy of the approach [5].

iii) Fault allocation technique based on synchronised phase voltage: - This paper proposes a novel fault location technique for distribution networks based on PMU synchro phasor voltages. Data gathered from as little as two PMUs can be utilised with the recommended methodology. The fault spot is identified using the bus admittance matrix in conjunction with the pre- and during-fault voltage phasors. The bus nearest to the fault is where it is initially gathered, and an iterative process is then employed to pinpoint the exact site of the fault. The technique's main innovative features are that it can identify the faulted branch instead of merely the closest bus, that it does not require current phasors, and that it does not require fault type classification. A real 134 bus feeder that has been modelled in ATP-EMTP is used to validate the performance of the suggested technique. The outcomes show a resilient performance in the face of various fault circumstances [6].

iv) Branch Fault Judgement Fault allocation technique based on matrices: - This work proposes a fault branch judgement matrix-based fault location method for transmission lines with multiple terminals. The fault travelling wave signal is broken down using Complete Ensemble Empirical Mode Decomposition with Adaptive Noise (CEEMDAN). Then, by creating a correlation-rearrangement entropy function, the IMFs sensitive components that can be used to determine the fault characteristics of the target signals are selected. The Teager Energy Operator (TEO) was used in conjunction with fault signals to enable accurate calibration of the fault signals' arrival time at the endpoint. In order to improve location accuracy by mitigating the effects of wave velocity and fault duration on the

location findings, this work proposes a two-terminal location approach based on the line mode component. This allows the fault branch judgement matrix to be used in conjunction with it to accurately locate multi-terminal transmission line faults. It has been shown that this approach yields fault detection results with a low error rate, excellent travelling wave head recognition, and precise fault branch judgement. It satisfies the criteria of real-world working scenarios and offers definite advantages over the existing multiterminal transmission line fault location algorithm [7].

v) Phasor measurement-based fault allocation technique: This study presents a novel approach to fault location on electrical transmission lines. This technique, which uses the phasor readings at one end of the line, makes use of the benefits of numerical relaying and digital technology, both of which are already available and may be used for offline analysis. This method only uses end of data and pinpoints the exact distance to the fault after one cycle from the start of the fault. The analysis for fast fault identification is assessed based on the travelling waves' representation by the wavelet modulus maxima. The kind, position, and instant of the defect can all be determined by the present criterion. The MATLAB/Simulink software was used to test the recommended approach. Different fault scenarios were simulated on a particular power system model by adjusting the fault type, fault resistance, fault location, and fault inception angle. The simulation results validate the validity of the proposed defective phase selection approach [8-14].

vi) PMU application-based fault analysis: - This work advances the field of fault site identification that improves grid dependability using an innovative approach. This article proposes a study-based approach that uses abnormal condition voltage and current frequency analysis for fault type classification and fault allocation. While this should always be the case, there are situations in which the frequency of the voltage and the current differ. This frequency variation is measured individually using the phasor measurement unit (PMU) block in the MATLAB/Simulink platform. The model for the PMU (PLL-based, positive sequence) block was IEEE Std C37.118.1-2011. Using this method, we measure the line voltage and current frequency variation using an installed PMU. Next, we

provide this measurement in characteristic form using the MATLAB/Simulink scoping tool, examine each characteristic separately, and suggest a course of action for fault location detection and fault type classification. The proposed method can distinguish between symmetrical and asymmetrical fault types and find faults on both the source and load sides. The suggested approach is tested using two MATLAB/Simulink models, and the outcomes are acceptable [15].

4. Comparative study

Sl. No.	Name of technique	Way of analysis	Remark
1	Impedance Based Fault Location Technique	Voltage sag calculation	Enable to fault location and fault types
2	Two Terminal Fault Combination Location Technique	Combine travelling wave and natural frequency fault location technique	Enable to identify fault location and fault types
3	LV Side Measurement Based Fault Allocation technique	Analysis of Characteristics voltage and current at LV side	Enable to identify suspected faulted path and faulty segment
4	Synchro phasor Voltage Based Fault Allocation Technique	Combining pre- and during-fault voltage phasors by bus admittance matrix	Enable to identify faulted branch
5	Fault Branch Judgment Matrix Based Fault Allocation Technique	Fault Branch Judgment Matrix	Fault detection results with low error rate
6	Phasor Measurement Based Fault Allocation Technique	Phasor data analysis at one end of line	Enable to locate fault, fault type and instant of fault
7	PMU Application Based Fault Analysis Technique	Voltage and current phasor frequency characteristics analysis	Enable to detect, allocate and identify fault nature of short circuit type fault

5. Conclusion

This paper presents the critical review of the fault detection, allocation and fault classification techniques/methodologies. The reviewed techniques

are developed in the recent years and can utilize in the dynamic power system. This paper also discusses the comparative study of the techniques".

References

1. S. N. Makhadmeh, A. T. Khader, M. A. A. -Betar, S. Naim, A. K. Abasi, and Z. A. A. Alyasseri, "Optimization methods for power scheduling problems in smart home: Survey," Renewable and Sustainable Energy Reviews, vol. 115, p. 109362, Nov. 2019.
2. G. Dileep, "A survey on smart grid technologies and applications," Renewable Energy, vol. 146, pp. 2589-2625, Feb. 2020.
3. Y. Menchafou, M. Zahri, M. Habibi, H. E. Markhi, "Extension of the Accurate Voltage-Sag Fault Location Method in Electrical Power Distribution Systems", Journal of Electrical Systems, Vol. 12, No.1, pp. 33-44, 2016.
4. M. Zhang, W. Zhao and P. Li, "Two-terminal fault combination location method on MMC-HVDC transmission lines based on ensemble empirical mode decomposition", Energy Reports, Vol. 9, pp. 987-995, 2023.
5. D. Zhang, W. Zhang, C. Wang and X. Xiao, "Fault Segment Location for MV Distribution System Based on the Characteristic Voltage of LV Side", Electronics, Vol. 12, pp. 1734-1755, 2023.
6. D.L.C. Crespo and M. Moreto, "New technique for fault location in distribution systems using sinchrophasor voltages", Electrical Power System Researchs, Vol. 212, No. 1, p. 108485,2022.
7. Y. Yang, Q. Zhang, M. Wang, X. Wang and E. Qi, "Fault Location Method of Multi-Terminal Transmission Line Based on Fault Branch Judgment Matrix", Applied Sciences, Vol. 13, No. 2, pp. 1174, 2023.
8. A. Rao J and B. Bogale, "Accurate Fault Location Technique on Power Transmission Lines with use of Phasor Measurements", IJERT, Vol. 4, No. 2, pp. 492-495, 2015.
9. S.P. Singh, S.P. Singh, "Optimal cost wide area measurement system incorporating communication infrastructure", IET Generation, Transmission & Distribution, vol. 11, pp. 28142821, 2017.
10. S.P. Singh and S.P. Singh, "Synchronized PMU Placement Incorporating Communication Infrastructure," 2016 7th India International Conference on Power Electronics (IICPE), pp. 1-5, 2016.

11. S.K. Bairwa and S.P. Singh, “PMU Placement Optimization Techniques for Complete Power System Observability: Literature Survey”, Journal of Chengdu university of Technology, vol. 26, pp. 1-11, 2021.
12. S.P. Singh and S.P. Singh, “Optimal PMU Placement in Power System Considering the Measurement Redundancy”, Advance in Electronic and Electric,Engineering, vol 4, pp. 593-598, 2014
13. S.P. Singh, A.K. Thakur, S.P. Singh, “PMU placement for maximum observability of power system under different contingencies”, Energy Procedia, vol.117, pp. 893-900, 2017.
14. S.P. Singh and S.P. Singh. "A multi-objective PMU placement method in power system via binary gravitational search algorithm", Electric Power Components and Systems, pp.18321845,2017
15. S.K. Bairwa and S.P. Singh, “Phasor measurement unit Application Based Fault Allocation and Fault Classification “International Journal of Advances in Applied Sciences, vol.12, pp.15-26, 2023.

Chapter - 32

An Overview of the Power Systems Studies' use of Machine Learning

Authors

Rituparna Mukherjee
Electrical Engineering, Swami Vivekananda University, Kolkata, West Bengal, India

Susmita Dhar Mukherjee
Electrical Engineering, Swami Vivekananda University, Kolkata, West Bengal, India

Abhishek Dhar
Electrical Engineering, Swami Vivekananda University, Kolkata, West Bengal, India

Promit Kumar Saha
Electrical Engineering, Swami Vivekananda University, Kolkata, West Bengal, India

Rituparna Mitra
Electrical Engineering, Swami Vivekananda University, Kolkata, West Bengal, India

Avik Dutta
Electrical Engineering, Swami Vivekananda University, Kolkata, West Bengal, India

Chapter - 32

An Overview of the Power Systems Studies' use of Machine Learning

Rituparna Mukherjee, Susmita Dhar Mukherjee, Abhishek Dhar, Promit Kumar Saha, Rituparna Mitra and Avik Dutta

Abstract

Due to the complexity of today's electric power networks originating from generation, transmission, and distribution stations, managing real-time data acquisition, supervisory control, and monitoring has produced a significant amount of complex data that requires further technical and mathematical analysis. This has led to the need for more accurate analysis and projections in power systems research, especially in scenarios including uncertainty, transience, or emergencies where human intervention is not an option. Reducing mistakes is essential for increasing overall performance. Therefore, it is now imperative to use extremely clever yet complex prediction systems. Machine Learning (ML) is a powerful tool that makes it possible to accurately predict future properties of data by using historical data. Machine Learning (ML) algorithms build a model mathematical or pictorial from input examples in order to produce data-driven predictions or judgements for the future. Big data and Machine Learning (ML) can be used to solve complicated data analytics problems and develop effective predictive systems. Several forecasting methods for electricity generation have been proposed, with the potential to anticipate required power at a rate similar to that of energy consumption. This work aims to review machine learning applications to power system studies. This article explored applications of machine learning methods in power systems investigations.

Keywords: Machine learning, big data, data analytic, power systems**.**

Introduction

Among the most intricate systems that humans have ever created is the electric power system. Among the elements that must be controlled for the network to operate well are economy and supply security. In a nutshell, a

power system's security refers to its ability to maintain supply continuity in the face of a variety of disruptions, including shifts in customer demand, internal malfunctions, and external disruptions like lightning and storms. Typically, security is attained by using two complimentary tactics [1]. Both emergency and preventive control. The former involves human operators keeping the system in a state that allows it to resist disruptions, whereas the latter operates automatically in the wake of a disturbance to reduce its consequences. "Since disturbances are intrinsically random, preventive control is essentially aimed at balancing the economic cost of normal operation against the risk of instability/insecurity. On the other hand, emergency control is aimed at reducing the severity of instabilities. Worthy of note is the fact that due to increased competitive pressure in the electricity industry and the possibilities offered by modern communication and computing technologies, the trend in power systems is to rely more on emergency control." Establishing suitable criteria that allow a real-time prediction of whether a system is losing stability or not is one of the issues in designing emergency controls. This means that proper real-time measurements must be chosen from a wide range of options, then combined to create detection criteria. The design approach for emergency controls mostly uses numerical simulation of the power system under various scenarios that could cause instability. Techniques like the Monte-Carlo sampling approach can be used to achieve screening of very large samples (several thousands) of large-scale simulations, providing massive data bases of simulation results. This is made possible by the continuously increasing amounts of computing power. Subsequently, these data can be utilised using automatic learning such as machine learning with a view to extracting useful information.

Literature review

[2] has suggested an effective machine learning-based big data forecasting method for the production of power. The study demonstrates that the amount of resources (coal, gas, nuclear, petroleum, oil, and renewable energy) required to generate electricity directly correlates with the amount of power produced. According to studies, forecasting electricity generation may be utilised to provide imprecise information about the demand for power and, likely, the need to raise the amount that must be imported from nearby nations. The necessity for accuracy makes the forecast difficult. When there is a lot of noise and volatility in the data sets, the prediction becomes much more difficult. Numerous forecasting techniques utilising various machine

learning algorithms have been put out to address forecasting issues related to power. Support vector regression (SVR), Grey Algorithm (GA), Fuzzy Neural Networks (FNN), and Grey Markov Model (GMM) are a few of these algorithms [2]. According to research, these models have produced very good forecasting outcomes. The forecast is challenging due of the requirement for precision. The prediction gets significantly more challenging when the data sets contain a lot of noise and volatility. Various machine learning algorithms have been employed in a multitude of forecasting methodologies to tackle power-related forecasting problems. Among these algorithms are Fuzzy Neural Networks (FNN), Grey Algorithm (GA), Grey Markov Model (GMM), and Support Vector Regression (SVR) [2]. Research indicates that these models have yielded very good forecasting results. Additionally, it was said that several methods, including frequency domain spectroscopy (FDS), dissolved gas analysis (DGA), and measurements of polarisation and depolarisation currents (PDC), have been employed by utilities for transformer diagnostics in recent years. Nonetheless, establishing a meaningful connection between measured data and the actual condition of transformer insulation remains extremely difficult. Self-Organising Maps (SOM) and Support Vector Machines (SVM), two machine learning (ML) algorithms created for the project, are used to diagnose transformer insulation systems and automatically analyse measurement data. The key advantages of these algorithms is their capability of acquiring the knowledge of underlying statistical dependency between archived data and the conditions of corresponding transformers and making use of such knowledge to assist in transformer insulation diagnosis. A proposal for ML-based power transformer lifespan prediction can be found in [4]. According to reports, energy companies find it crucial to anticipate the remaining life of high voltage power transformers since it helps them prepare for capital expenditures and maintenance. Both the examination and lifetime prediction of the transformers were done using machine learning techniques. It is shown that transformer lifespan estimation was improved by combining machine learning models with experimental models. To forecast the lifespan distribution of each transformer, a parametric lifetime model is employed. A statistical process is created to determine how long each individual transformer that is currently in operation will last. Before damage to the transformers may happen, the power transformer loss values are sent to the end users via email or any other form of communication by use of machine learning algorithms. By doing this, end users can be informed about how

long the transformer will last and take preventative measures, such as performing regular maintenance, to prevent the transformer from failing. In [5], the design of various intelligent agents for power system control is examined using a novel kind of reinforcement learning algorithm called "fitted Q iteration." The algorithm's primary feature is its formulation of the reinforcement learning issue as a series of conventional supervised learning problems. "Fitted Q iteration" is said to have the ability to handle power system control in the actual world. There are two issues that can be exploited to get around the issues with using the "fitted Q algorithm." A approach designed to operate a Thyristor-Controlled Series Capacitor (TCSC) deployed in a 4-machine power system in real-time yields encouraging simulation results. According to [6], contemporary power systems that incorporate more renewable energy generation and have a higher degree of demand side participation are dealing with a growing number of complications and uncertainties. It is also stated that better methods for system modelling, assessment, and decision-making are necessary for the grid to operate reliably. With the deployment of smart meters and cutting-edge sensing technology, system operators may now access fine-grained electrical data, however there is an urgent need for efficient and near-real time algorithms to analyse and make better use of these available data. Recent advances in machine learning (ML) algorithms especially the giant leaps in deep learning makes ML a good tool for solving a series of data driven problems in power systems. For instance, ML methods such as Recurrent Neural Networks (RNN) can find straight forward applications in wind, solar power and building load forecasting. ML has been applied to power grid outage detection. High Voltage Alternating Current (HVAC) control and grid protection policy formulation problems can also solved using ML approaches. In [7], it is reported that increased use of renewable energy liberalization of energy markets and most importantly the integration of various monitoring, measuring, and communication infrastructure into modern power system networks offers the opportunity for building a resilient and efficient power grid network at various voltage levels. The numerous risks of instability and insecurity to the intricate networks, such as cyberattacks, voltage instability, and disruptions in power quality (PQ), are another cause for concern. Additionally, machine learning (ML) techniques including support vector machines (SVMs), decision trees (DTs), and artificial neural networks (ANNs) have been applied to safe and stable power system operations to make effective decisions and control actions. The most

recent research on machine learning technologies (MLTs) for power system security and stability, particularly in cyberattack detection, PQ disturbance investigations, and dynamic security assessment studies, is thoroughly reviewed in this paper. The study's goal is to draw attention to the approaches, successes, and above all the shortcomings in the data collection, test systems, and classifier design that were used. "A brief review of Reinforcement Learning (RL) and Deep Reinforcement Learning (DRL) approaches to transient stability assessment is also presented. The research gap highlighted states that despite the enormous accomplishments in terms of power system investigations in the areas of security and stability studies, a number of challenges still remain unresolved. The prediction and detection accuracy of MLTs are known to depend majorly on the quality and quantity of the data set and test systems employed. However, sometimes due to the non-availability and inadequacy of realistic power systems, data from real power stations and field devices, scholars and researchers have been restricted to the use of simulated data sets and the development of scalable test beds which have shown inconsistency in predictions and classification." The adjustment of the parameters is another significant feature unique to ML applications, in addition to the quantity of input data sets. Because of the complex processes involved in adjusting the parameters to produce the intended outcomes, MLT techniques necessitate a high degree of expert participation. Furthermore, MLT techniques can occasionally need a lot of time. Additionally, the majority of publications in the literature often use the assumption that PMU data are accurate, comprehensive, and accessible online. In actual use, jamming, malfunctions, and attacks could prevent the measurements from being available all the time. Ultimately, it was recommended that future research on MLT-based approaches to power system security and stability studies concentrate on thorough validation of the methods utilising sizable test systems with attributes that are comparable to modern power systems. It was discovered in [8] that in order to guard against problems on power transmission lines, contemporary power systems needed to be monitored in real time and controlled quickly. As the Smart Grid takes shape, installing high-quality sensors—such as smart meters, remote terminal units, and phasor measurement units—tends to produce a sizable volume of heterogeneous data that is needed for the grid's operational control and performance analysis. From a computational perspective, conventional time-domain techniques tended to be wasteful and could not match the requirements of real-time applications. Nevertheless, machine

learning (ML) methods allow the data to be learnt without being explicitly programmed, and after it is exposed to fresh data, it can react on its own. Artificial Neural Networks (ANNs), Decision Trees (DTs), Deep Learning Models (DLM), and other machine learning techniques can yield fascinating insights on power system safety. The study suggested using machine learning to categorise and identify issues in power systems. A list of ML approaches for fault classification was provided by the study. They are listed in order;

Decision trees, Bayesian learners, Support Vector Machines, Sequential Minimal Optimisation, Logistic Regression, and K. Nearest Neighbour.

Four potent and well-liked machine learning techniques (Bagging Classifier, Boosting Classifier, Radial Basis Function Classifier (RBF), and Naïve Bayes Classifier (NBC)) were established in [9] for the purpose of locating and recognising defects over a 600 km power transmission line. Eleven (11) defects are determined to be identified, projected, and located in the article. Given that high prediction accuracy is attained, the experiment's results point to the potential applications of RBF, Bagging Classifier, and NBC approaches for fault type prediction. The obtained prediction findings for the fault site prediction experiment were not as precise as the fault type prediction results. However, the maximum prediction accuracy was attained by the RBF, NBC, and Bagging Classifier. Lastly, it was mentioned that transmission line defects may be found using machine learning approaches. To locate faults precisely, location prediction accuracy results might need to be enhanced. A thorough analysis of the identification, categorisation, and positioning of faults in transmission lines is provided in [10]. An overview of feature extraction techniques was given before techniques for fault detection, classification, and location were covered. The foundation for dimensionality reduction strategies, several transforms, and defect identification algorithms were covered. There was also discussion of recently created concepts and how they compared to several notable fault detection aspects. It is demonstrated that ML techniques are frequently used by researchers to categorise different failure types. But for fault classification, it's also advised to use Fuzzy Inference System (FIS), Artificial Neural Networks (ANNs), Decision Trees (DT), Convolutional Neural Networks (CNN), and Restricted Boltzmann Machine Learning (RBML), which are Deep Learning-based algorithms. The use of artificial intelligence (AI) in conjunction with fault location detection algorithms was covered. Because transmission systems require more processing and communication than ever before, deep learning

techniques were suggested for fault location techniques in the future. Here's a rewritten version with reduced similarity to the original: A study [11] explored the use of Artificial Neural Networks (ANNs) for identifying, categorizing, and pinpointing faults in electrical systems. The research employed forward-propagating neural networks trained using backpropagation techniques. To verify the proposed fault detection method, researchers simulated a WSCC 9-bus system using MATLAB/SIMULINK. The findings demonstrated ANNs' capability to effectively detect, classify, and locate faults, even in scenarios they weren't specifically trained for. The study revealed that different fault types require varying network architectures. For instance, single-phase ground faults could be addressed with minimal hidden layer neurons, while more complex faults necessitated larger networks. Specifically, the research found that detecting and locating single-phase ground faults required only five neurons in the hidden layer. Two-phase and two-phase-to-ground faults needed at least 10 hidden neurons, while three-phase faults called for 30-35 hidden neurons. Interestingly, adding more neurons beyond these numbers did not enhance performance. In [12] a related study, researchers leveraged deep learning frameworks to automate disturbance classification in power systems. This approach was motivated by the vast amounts of data available from modern grid monitoring systems. The team utilized measurements from multiple Phasor Measurement Units (PMUs) installed in low-voltage areas of an interconnected network. From this data, they extracted representative patterns to train a classifier capable of recognizing various system disturbances. Specifically, MultiLayer Perceptrons (MLP), Deep Belief Networks, and Convolutional Neural Networks (CNN) are the techniques used; CNN has proven to be more accurate in classifying data than the other two. Furthermore, these architectures were put into practice in the CPU and the GPU in order to determine the ensuing speed increases. The use of artificial neural networks (ANNs) for fault detection and classification on a three-phase transmission line system is described in [13]. The technique developed uses three-phase currents and three-phase voltages as neural network inputs. According to their respective pre-fault values, the inputs were normalised. The findings were limited to line-to-ground problems. The back propagation neural network architecture was employed in the ANNs investigations that were used in this paper. The resulting simulation results demonstrate that all of the suggested neural networks performed satisfactorily and are practically implementable. The study emphasises how crucial it is to select the ideal ANN configuration in order to obtain the best

results. The research leads to the following significant conclusions, which are;

- "ANNs are a reliable and effective method for an electrical power system transmission line fault classification and detection in view of the increasing dynamic connectivity of the modern electrical power transmission systems."
- "The performance of an ANN should be analysed properly. (In particular, a neural network structure and learning algorithm should be analysed properly before choosing it for a practical application)."
- "Back propagation neural networks deliver good performance when trained with large training data sets which is easily available in power systems."

"In [14], a method based on a combination of wavelet singular value and Fuzzy Logic (FL) is presented for fault detection and fault classification in power transmission systems. The results show that the proposed indices for FL are sensitive to variations. The method is robust to parameter variations such as fault type, fault inception location, fault resistance and power angle and can properly detect faults. The proposed algorithm has proven to be a convenient and rapid method for fault detection and fault classification in different conditions and is able to detect and classify faults and determine a healthy phase from a faulty phase in less than 10ms after fault inception. In [15], a reliable scheme for the detection, classification, and location of faults on transmission lines is developed. The scheme combines the feature extraction capability of the discrete wave transform (DWT) and the intelligent classification capability of the Adaptive Neuro Fuzzy Inference System (ANFIS). The developed DWT-ANFIS model is tested and the results compared with Impedance-ANFIS model. Faults are detected within 8ms that is less than one complete cycle from fault inception to prevent equipment damage and prolonged power outage. In [16], a wavelet transform-based approach to detect and classify different shunt faults that may occur in transmission lines is presented. The algorithm is based mainly on calculating the RMS values of the wavelet coefficients of current signals at both ends of the transmission lines over a moving window length of half cycle. The current signals are analysed with "dB4" wavelet to obtain detail coefficients and compared with threshold values to detect and classify the faults. To illustrate the effectiveness of the proposed technique, extensive simulations using PSCAD/EMTDC and MATLAB have been carried out for different

types faults considering wide variations of resistances, inception angle, and loading levels. The study proposes that the techniques investigated are well suited for implementation in digital distance protection schemes. In [17], an endeavour aimed at the automation of power system fault identification using information conveyed by the wavelet analysis of power system transients is proposed. Probabilistic Neural Network (PNN) is also used in the study. The focus of the study is on the identification simple power system faults. Also performed was a wavelet transform of the transient disturbance caused as a result of the occurrence of a fault. The detail coefficient for each type of simple fault is characteristic in nature. PNN was used in distinguishing the detailed coefficients and hence the fault. The application of wavelet transforms to determine the type of fault and its automation incorporating PNN could achieve an accuracy of 100% for all type of faults. Back propagation algorithm was limited in distinguishing the entire phase-to-ground and double line to-ground fault. In [18], the study says that transmission line relaying involves three major tasks."

- Fault detection
- Fault classification
- Fault location

To de-energise the malfunctioning line, these three actions need to be completed as quickly and precisely as feasible. In light of these and other factors, the study suggested a unique technique for utilising oscillographic data to detect and classify transmission line faults. A set of criteria derived from the current waveform analysis in the time and wavelet domains were used to determine the fault identification and clearing. The technique can distinguish faults from other power quality disruptions, like oscillatory transients and voltage sags, which are frequent in power system operation. An artificial neural network (ANN) uses the identification of voltage and current waveform patterns in the time domain to classify the defect. Excellent results were reportedly obtained when the method was applied to genuine oscillographic data from a Brazilian power company for the purpose of defect identification and categorisation. The goal of the study in [19] was to create a single artificial neural network (ANN) that could identify and categorise faults on the 33 kV electric power transmission lines in Nigeria. To create a fault detector-classifier, the study used a feed-forward artificial neural network using a backpropagation algorithm. The effectiveness of the developed intelligent systems for fault detection and classification on 33 kV

lines has been demonstrated by simulation results. The confusion matrix and mean square error (MSE) are used to assess the detector-classifier's performance. The system's accuracy of 95.7% and respectable MSE of 0.00004279 demonstrate the developed intelligent system's satisfactory performance. The developed system in this work is deemed to be superior than other systems in the literature regarding Nigeria's transmission lines. The study conducted in [20] focused on the potential application of convolutional neural networks (CNNs) and deep learning architecture for real-time power system fault classification. The effort focused solely on fault classification rather than localisation, with the goal of categorising real-time power system voltage signal samples and identifying whether they were in a faulty or non-failed state. A straightforward two-bus power system with a three-phase balanced load is simulated to provide the data. While the fault occurs at half of the line length between the two buses, the voltage signal is measured at the beginning of the line between the two buses. Using dB4 daubechies mother wavelets, the wavelet transform is utilised in the first step to extract the fault harmonics. At decomposition level 4, a fixed-size sample window is slid over the wavelet information, which appears to be a good option. The produced training samples are fed into the convolutional neural network (CNN) for the learning process after normalisation. In order to ultimately categorise test sample data, the CNN is trained using both defective and non-faulted samples to identify fault characteristics of the power system. "In conclusion, it was shown that CNNs could successfully learn power system fault's features and classify those correctly. For certain training scenarios (only faulted test samples) as per phase testing accuracy of over 85% is achieved. This has been validated by a simulation of a two-bus power system with balanced load. In [21], it was reported that in consideration of machine learning applications, it has become easier to handle complex power system challenges. The traditional techniques are not computationally promising solutions since they are limited in capacity to manage the massive amounts of data (including chunks of heterogeneous datasets) coming from measurement units such as smart meters, and phasors measurement units (PMUs). The study said there was in existence several advanced, efficient, and intelligent learning algorithms that have been developed to improve the accuracy of solutions to many real-world problems in a diversity of areas such as voltage stability, power flow management, rotor system diagnosis to mention a few. The study indicates that supervised machine learning classifications are more in use compared to other methods. What this means is that classification algorithms yield more benefits to problems than others.

The study concludes by saying that it can be inferred that by applying machine learning to electrical engineering problems, difficult issues are not only simplified but also results secured are also reliable and precise. In [22], Machine learning (ML) techniques for power system security assessment is presented. It is reported that modern electricity grids continue to be vulnerable to large scale blackouts. It is also reported that as all states leading to large scale blackouts are unique, there is no algorithm to identify pre-emergency states. Moreover, numerical conventional methods are computationally expensive which makes it difficult for them to be used for online security assessment. Machine learning techniques with their pattern recognition, learning capabilities, and high-speed identification of potential breach of security boundaries can offer an alternative approach. The study put forth that during the last 10 years, events in the North American continent, in Europe, and in Asia has clearly demonstrated an increasing likelihood of large blackouts. This indicates that the security monitoring and control of power systems need to be improved. The paper presents a novel method for online security assessment using machine learning techniques. Multiple machine learning techniques such as Artificial Neural Networks (ANNs), Support Vector Machines (SVM), Decision Trees (DTs) are first trained online using the resampling cross validation method. Resampling the training samples allows to know when a poor choice of values of the machine learning tuning parameters is being made. The best model of the ML technique is selected based on its performance. For the online application, the final of the best of the best ML is used as the candidate technique with the best performance. If required the final ML is checked and updated in order to account for new changing system states as accurately as possible. The results obtained in the work showed that the proposed approach can identify potentially dangerous states with high accuracy, and if required the final ML model can produce an alarm for triggering emergency and protection systems. In [23], an exploration of how Reinforcement Learning (RL) as applied to the control of power systems is presented. Also presented is a description of some challenges in power system control as well as how some of those challenges can be overcome using RL techniques. The difficulties associated with application of RL methods to power system control as well as the strategies to overcome them are also presented. Two RL modes are considered."

- The online mode in which interaction occurs with the real power system.

- The offline mode in which interaction occurs with a simulation of the model of the real power system.

Two case studies made on a 4-machine power system model are presented where in the first case, the design of a dynamic brake controller by means of RL algorithms used in the offline mode is considered. The second case concerns RL methods used in the online mode when applied to control a thyristor-controlled series capacitor aimed at damping power system oscillations. The RL methods can reveal themselves to be an interesting tool for power system agents' design for reasons enumerated;

- RL methods do not make any strong assumptions on the system dynamics. In particular they can cope with partial information, nonlinear and stochastic behaviours. They can therefore be applied to the design of many practical types of control schemes.
- This method learns closed loop control laws which is ascertained to be robust. This aspect is important notably when the real power system is faced with situations that were not accounted for in the simulation model.
- RL methods open avenues to adaptive control since the RL driven agents learn continuously and can adapt to changing operating conditions.
- The method can be used in combination with traditional control methods to improve performances.

In [24], it was presented that power system protection includes the process of identifying and correcting faults before fault currents cause damage to utility equipment or customer property. In distribution systems where the number of measurements is increasing, there is an opportunity to improve upon fault classification techniques. Fault classification using machine learning (ML) techniques and quarter-cycle signatures is presented. Separate voltage and current-based feature vectors are defined using multi-resolution analysis are input to a two-stage classifier. The classifier is trained and tested on experimental fault data using a Reconfigurable Distribution Automation and Control software/hardware laboratory. Results show Non-linear and even Non-contiguous decision regions on a fault plane using a phase voltage based feature. An accurate classifier for determining the grounding status of multiphase faults using a neutral current based feature. In [25], the focus is on the concept of using reinforcement learning to control the

power system's unit commitment and economic dispatch problem. The idea of reinforcement learning strives to present an ever-optimal system even when there are load fluctuations. This is done by training the agent (system) thereby enriching its knowledge base which ensures that even without manual intervention, all the available resources are judiciously used. Also, the agent learns to reach the long-term objective of minimizing cost by autonomous optimization. A model-free reinforcement learning method called "Q Learning" is used to find the cost at various loadings and is compared with the conventional priority list method and the performance improvement due to Q learning is proved. For future directions, it was presented that single agent reinforcement learning can be extended to multi-agent reinforcement learning to accommodate other types of renewable energy resources such as solar and wind along with thermal units. In [26], an active Machine Learning (ML) technique for monitoring the voltage stability in transmission systems is presented. It is shown that ML algorithms may be used to supplement the traditional simulation approach. However, they suffer from difficulties associated with online ML model update and offline training data preparation. An active learning solution to enhance existing ML applications by actively interacting with the online prediction and offline training processes is presented. The method identifies operating points where ML predictions based on power system measurements contradict with actual system conditions. By creating the training set around the identified operating points, it is possible to improve the capability of ML tools to predict future power system states. The method also accelerates the offline training process by reducing the amounts of simulations on a detailed power system model around operating points where correct predictions are made. Experiments show a significant advantage in relation to the training time, prediction time, and number of measurements that need to be queried to attain high prediction accuracy. In [27], a summary of artificial intelligence (AI) and its increasingly widespread application control is presented. Compared with traditional technologies, AI technologies possess obvious advantages. The application of intelligent technology in electrical automation control systems can reflect the basic characteristics of high precision, high efficiency, and high coordination. The application of AI while achieving automatic control can greatly improve the operating efficiency quality of the control system. Intelligent control can also realize optimal allocation of resources, reduce resource cost investment, improve the economic benefits of related companies, and promote the sustainable

development of a nation's electrical industry. In [28], the study addresses the on-going work of the application of machine learning (ML) to the dynamic security assessment of power systems. It lists several methods which have been applied to the Greek power system. These methods include;

- Offline Supervised Learning (Radial Basis Function Neural Network (RBFNN), Support Vector Machine (SVM), Decision Trees (DT)).
- Offline Unsupervised Learning (Self-Organizing Maps (SOM))
- Online Supervised Learning (Probabilistic Neural Network (PNN))

The results from the application of the ML methods show the accuracy and versatility of the methods. RBFNN and SVM perform not only classification of the system states but also regression and gives an estimation of the security criterion (voltage and/or frequency) value. This is very important as it can be used as a measure of the security margin of the system. Decision Trees on the other hand provide explicit rules to the system operator while inverse reading of the rules can also establish load shedding schemes when the safety of the system is jeopardized. The advantage of the SOM in comparison to the offline supervised learning methods is that its construction is independent of the security criterion applied. This means that when the classification criterion is change, the only modification required is the straightforward recalculation of the security indices for each of the map's node. On the contrary, a change in the classification would require the reconstruction of any of the offline supervised learning methods. The advantage of the online learning method such as PNN is that it can deal with changes in the structure of the power system without the need to completely retrain the PNN. In [29], it is presented that with increasing complexity, uncertainty, and data dimensions in power systems, conventional methods often meet with bottlenecks when trying to solve decision and control problems. Data-driven methods aimed at solving the decision and control problems are currently being extensively studied. Deep Reinforcement Learning (DRL) is one of such data driven methods and is regarded as real artificial intelligence (AI). DRL is a combination of deep learning (DL) and reinforcement learning (RL). DRL has achieved rapid development in solving sequential decision-making problems around theoretical, methodological, and experimental fields. In particular, DL obtains an objects attributes or characteristics from the environment while RL makes decisions with regards to the control strategies according to the information. Therefore, DRL can solve problems in large, high dimensional states, and action spaces.

As power systems evolve, to reflect the smart grid, there is in existence new challenges such as the integration of renewable energy and liberalization of the electricity market which offer difficulties to traditional techniques trying to solve problems in these areas. AI methods such as DRL can solve problems arising from these. In [30], a new control architecture for future power distribution system protective relay setting is envisioned. With increased penetration of distributed energy resources, at the end-user level, it has been recognized as a key engineering challenge to redesign the protective relays in the future distribution system. Conceptually, these protective relays are the discrete ON/OFF control devices at the end of each branch and node in a power network. The key technical difficulty lies in how to set up the relay control logic so that the protection could successfully differentiate heavy loads and faulty operating conditions. The study proposes a new nested reinforcement learning approach to take advantage of the structural properties of distribution networks and develop a new set of training methods for tuning protective relays.

Machine learning

"Machine learning is the study of computer algorithms that improve automatically with time through experience by the use of data obtained for the study of interest. These algorithms are used to build a model (mathematical, pictorial etc.) based on sample data known as "training data" in order to make predictions or decisions without being explicitly programmed to do so (i.e. it can easily be adopted to solve a particular task). Machine learning involves computer learning from data provided so that they can carry out certain specific tasks. For simple task assigned to computers, it is possible to programme algorithms notifying the machine on how to execute all steps required to solve the problems at hand; on the computer's part no learning is needed. For more advanced tasks, it can be challenging for a human to manually create the needed algorithms. In practice, it can turn out to be more effective to help the machine develop its own algorithm rather than having human programmers specify every needed step. Machine learning (ML) is a subfield of artificial intelligence which deals with the study of algorithms which learn from past experiences. The goal of machine learning is to give the attribute of intelligence to computers and ultimately machines. Machine learning involves computers discovering how to carry out tasks without being explicitly programmed to do so. It involves computers learning from data provided so that they can carry out certain task. For simple tasks assigned to computers, it is possible to

programme algorithms telling the machines how to solve the problem at hand. On the computer's part no learning is needed. For more advanced tasks, it can be challenging for a human designer to manually create the needed algorithms. In practice it can be more effective to help the machine develop its own algorithm rather than having human programmers specify every needed step. Machine learning has overlapping similarities with expert systems in which a computer system emulates the decision-making ability of a human expert. Expert systems are designed to solve complex problems by reasoning through bodies of knowledge. An expert system is composed of two subsystems: the inference engine and the knowledge base. The knowledge base represents facts and rules. The inference engine applies the rules to the known facts to determine new facts."

a) Classification of machine learning algorithms

Machine learning techniques are conventionally divided into three broad categories depending on the nature of the "signal" of "feedback" available to the learning system. These techniques are;

b) Supervised learning

Supervised learning algorithms build a mathematical model of a set of data that contains both the input and the desired outputs given by a "teacher". The data is known as the training data, and consists of a set of training examples [31]. Each training example has one or more inputs and the desired outputs, also known as the "supervisory" signal. In the mathematical model, each training example is represented by an array or vector, sometimes called a feature vector, and the training data is represented by a matrix. Through iterative optimization of an objective function, supervised learning algorithms learn a function that can be used to predict the output associated with new inputs. An optimal function will allow the algorithm to correctly determine the outputs for inputs that were not part of the training data. An algorithm that improves the accuracy of its output or predictions over time is said to have learned to perform that task. Types of supervised learning algorithms include active learning, classification, and regression.

c) Unsupervised learning

Unsupervised learning algorithms take a set of data that contains only inputs and find structure in the data like grouping or clustering of data points. The algorithms therefore learn from test data that has not been labelled, classified, or categorized. Instead of responding to feedback, unsupervised learning algorithms identify commonalities in the data and

react based on the presence or absence of such commonalities in each new piece of data.

d) Semi-supervised learning

Semi-supervised learning falls between unsupervised (without any labelled training data) and supervised learning (with completely labelled training data). It has been found that unlabelled when used in conjunction with a small amount of labelled data, can produce a considerable improvement in learning accuracy.

e) Reinforcement learning

This is an area of ML concerned with how software agents ought to take actions in an environment so as to maximize some notion of cumulative reward. In ML, the environment is typically represented as a Markov Decision Process (MDP). Reinforcement learning algorithms do not assume knowledge of an exact mathematical model of the MDP and are used when exact models are infeasible.

f) Models

These models provide a mathematical framework for learning. A model is human derived and is based on human observation and experiences. Figure 1 shows a single neuron model. Different models have different mathematical and pictorial representations.

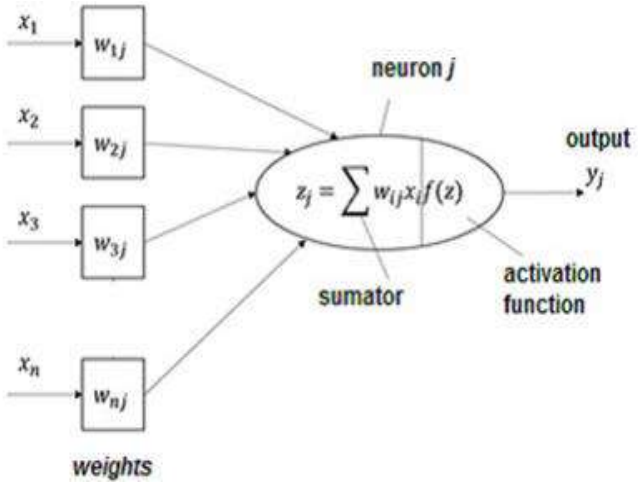


Figure 1: Model of a single neuron [11]

Performing ML endeavour involves creating a model, which is trained on some training data and then can process additional data to make predictions. Some of these models include;

- Artificial neural network: it is also called neural networks (NNs) as it is based on nodes connections that models the neurons used to transmit signals to other neurons formed in the same manner. This signal is informed of real numbers as input and the output is then computed as a non-linear function summing up its inputs.
- Decision trees: it is a tool that supports decision making using the model that tree-like in nature.it considers consequences of taking certain decision while putting certain factors such as cost of resources and utility as well as possibilities of outcome of events and summarily shows an algorithm that gives statements that are conditional.
- Support vector machine: This is a kind of computer algorithm that learns using examples to assign labels to objects and it has been successfully applied in solving some technical problems in the areas of power systems studies [32]. In other words, it can be defined as a formulated mathematical algorithm for maximizing a specific mathematical expression as a function of a given set of collected data. Sequential Minimal Optimization is an optimization tool for training support vector machine which requires a very large optimization programme.
- Regression analysis: This consist of a set of machine learning methods that enables prediction of an output variable based on the values of one or more predictor variable(s). It is termed machine learning because its task is to give an estimated value based on certain predictive features. It uses test sets to validate its accuracy of prediction.
- Bayesian network: It is defined as a probabilistic graphical model consisting of two parts (structure and parameters). The structure part is known as the directed acyclic graph for dependent and independent conditional expressions. Probability is used to represent all uncertainties within the model and it used to obtain the posterior probabilities based as well as additional recent information.

- Genetic algorithms: It is a stochastic search algorithm that has found vast applications in the area of ML to power systems studies.it has cross over, mutation and fitness selection as its three components.it is used to obtain solutions required for optimization and search problems through inspiration obtained biologically with respect to these three components.
- Deep learning: This is a sub-set of ML and it is mathematically a complex improvement of ML algorithm.it analyse data logically and then draws intelligent conclusion the way humans will make. This is achieved through supervised or unsupervised training pattern and it uses layer structure to achieve this.it is worthy of note that with the current high cloud computing and transfer learning ability of ML, the training time is now very fast and more accurate.
- K. Nearest Neighbour (KNN): It is used for analysing large data; thus, it can accommodate very large training data. In KNN, a set of variables characterise each data point.it is an algorithm developed to store all cases of existing variables and then make appropriate classification based on similarities of the previous cases. Its advantage is that no assumption is about the data used, thus making the final results obtained to be very accurate.

Conclusion

The applications of ML to power system studies are increasing as the size and complexities of the system continue to expand. This review as considered various ML strategies as they apply to power systems studies including the critical identification of their knowledge gaps as investigated by each researcher. Moreso, ML techniques models and algorithms as well as their applications to various sectors of the power network (generation, transmission and distribution) were reviewed. Various subdivisions of ML which include supervised, unsupervised and reinforcement learning have been discussed alongside their associated algorithms as applied to power system problems. Furthermore, advantages and disadvantages of some ML techniques were also considered. Various models used to achieve ML was also considered. These include Artificial Neural Network (ANN), Decision Tree, Support Vector Machine, Regression Analysis, Bayesian Network (BN), Genetic Algorithm (GA), Deep Learning (DL) and the K-Nearest neighbour (KNN)."

References

1. P Geurts, L Wehenkel. Early Prediction of Electric Power System Blackout by Temporal Machine Learning, A.A.A.I Technical Report, WS-98-07. 1998.
2. MA Rahman, A Ismailpour, J Zhao. Machine Learning with Big Data: An Efficient Electricity Generation Forecasting System, Published in Big Data Research. 2016; 5: 9-15.
3. H Ma, TK Saha, C Ekanayake. Machine Learning Techniques for Power Transformer Diagnosis. 1-6.
4. G Pooja, M Preethi, B Jeevabavatharani, IG Hemanandhini. Power Transformer Lifetime Prediction Through Machine Learning, Published in the International Journal of Creative Research Thoughts. 2020; 8(7): 1667-1670.
5. D Ernst, M Glavic, P Geurts, L Wehenkel. Approximate Value Iteration in the Reinforcement Learning Context: Application to Electric Power System Control", Published in the International Journal of Emerging Electric Power System. 2005; 3(1): 1066, 1-35.
6. Y Chen, Y Tan, D Deka. Is Machine Learning in Power Systems Vulnerable?", Published in the 2018 IEEE International Conference on Communications, Control and Computing Technologies for Smart Grids. 2018.
7. OA Alimi, K Ouhada, AM Abu-Mahfouz. A Review of Machine Learning Approaches to Power System Security and Stability, Published in IEEE Access. 8: 113512-113531.
8. D Baskar, P Selvam. Machine Learning Framework For Power System Fault Detection and Classification, Published in the International Journal of Scientific and Technological Research. 2020; 9(2).
9. AN Hassan, PSP Eboule, B Twala. The Use of Machine Learning Techniques to Classify Power Transmission Line Fault Types and Location", Published in the 2017 International Conference on Optimization of Electrical and Electronics Equipment. 2017; 221-226.
10. A Raza, A Benrabah, T Alquthami, M Akmal. A Review of Fault Diagnosing Methods in Power Transmission Systems", Published in the Journal of applied Sciences. 2020; 10: 1-27.

11. A Karic, T Konjic, A Jahic. Power System Fault Detection, Classification, and Location Using Artificial Neural Networks", Published in the International Symposium on Advanced Electrical Power Systems.
12. PEA Cardoso. Deep Learning Applied to PMU Data in Power Systems, Master of Engineering Thesis Submitted to the Department of Electrical & Computer Engineering, Faculty of Engineering, University of Porto.
13. M Jamil, SK Karma, R Singh. Fault Detection and Classification in Electrical Power Transmission System Using Artificial Neural Network", Published in Springer Open Journal. 2015; 1-13.
14. M Nayeripour, AH Rajaei, MM Ghanbarian, M Deghani. Fault Detection and Classification in Transmission Lines Based on a Combination of Wavelet Singular Values and Fuzzy Logic", Published in the Cumhuriyet University Faculty of Science, Science Journal. 2015; 36: 69-82.
15. FO Ogban, KM Udofia, CN Kalu. Fault Detection, Classification and Location on 132 kV Transmission Line Based on DWT and ANFIS", Published in the Journal of Multidisciplinary Engineering Science and Technology. 2020; 7(6): 12367-12376.
16. S Chakraborty, S Singh, A Bhalla, P Saxena, R Padarla. Wavelet Transform Based Fault Detection and Classification in Transmission Lines, Published in the International Journal of Research in Engineering and Applied Sciences. 2012; 2(5): 67-74.
17. KH Kashyap, UJ Shenoy. Classification of Power System Faults Using Wavelet Transforms and Probabilistic Neural Networks.
18. KM Silva, BA Silva, NSD Brito. Fault Detection and Classification in Transmission Lines Based on Wavelet Transform and Artificial Neural Networks, Published in the IEEE Transactions on Power Delivery. 21(4).
19. PO Mbamaluikem, AA Awelewa, IA Samuel. An Artificial Neural Network-Based Intelligent Fault Classification System for the 33 kV Nigerian Transmission Line, Published in the International Journal of Applied Engineering Research. 2018; 13(2): 1274-1285.
20. F Rudin, L Guo-Jie, K Wang. An Algorithm for Power System Fault Analysis Based on Convolutional Deep Learning Neural Networks,

Published in the International Journal of All Research Education and Scientific Methods. 2017; 5(9): 11-18.

21. SM Miraftabzadeh, M Pasetti. A Survey of Machine Learning Applications for Power System Analytics. 2019.
22. NV Tomin, VG Kurbatsky, DN Sidorov, AV Zhukov. Machine Learning Techniques for Power System Security Assessment, Published in IFAC-Papers Online. 2016; 49-12: 445-450.
23. D Ernst, M Glavic, L Wehenkel. Power System Stability Control: Reinforcement Learning Framework, Published in the IEEE Transaction on Power Systems. 2004; 1-9.
24. NS Coleman, C Schegan, KN Miu. A Study of Power Distribution System Fault Classification with Machine Learning Techniques, Proceedings at the 2015 North American Power Symposium. 2015.
25. L Raju, RS Milton, S Suresh, S Sankar. Reinforcement Learning in Adaptive Control of Power System Generation", Published in Procedia Computer Science. 2015; 46: 202-209.
26. V Mulbasa, C Zheng, C Po-Chen, T Popovic, M Kezunovic. Voltage Stability Prediction Using Active Machine Learning. Proceedings from the IEEE transactions on Smart Grids. 2017; 1-8.
27. C Jiang, X Xiong, T Zhu, J Cao, J Yu. Research on Application of Artificial Inteligence Technology in Electrical Automation Control, Published in the Journal of Physics. 2020; 1-6.
28. EM Voumvoulakis, AE Gavoyiannis, ND Hatziargyriou. Application of Machine Learning on Power System Dynamic Security Assessment, Proceedings from 14th International Conference on Intelligent System Applications to Power System, ISAP. 2007; 118-123.
29. Z Zhang, D Zhang, RC Qiu. Deep Reinforcement Learning for Power System Applications: An Overview, Published in the CSEE Journal of Power and Energy Systems. 2020; 6(1): 213-225.
30. D Wu, Y Zheng, Q Qian. Nested Reinforcement Learning-Based Control for Protective Relays in Power Distribution Systems, Proceedings from the 2019 IEEE 58th Conference on Decision and Control. 2019; 1925-1930.
31. https://en.wikipedia.org/wiki/Machine_learning.
32. WS. Noble What is a Support Vector Machine, National Publishing Group. 2006.

Chapter - 33

Detection of Transmission Line Faults: A Review

Authors

Rituparna Mukherjee
Electrical Engineering, Swami Vivekananda University,
Kolkata, West Bengal, India

Susmita Dhar Mukherjee
Electrical Engineering, Swami Vivekananda University,
Kolkata, West Bengal, India

Abhishek Dhar
Electrical Engineering, Swami Vivekananda University,
Kolkata, West Bengal, India

Promit Kumar Saha
Electrical Engineering, Swami Vivekananda University,
Kolkata, West Bengal, India

Rituparna Mitra
Electrical Engineering, Swami Vivekananda University,
Kolkata, West Bengal, India

Titas Kumar Nag
Electrical Engineering, Swami Vivekananda University,
Kolkata, West Bengal, India

Chapter - 33

Detection of Transmission Line Faults: A Review

Rituparna Mukherjee, Susmita Dhar Mukherjee, Abhishek Dhar, Promit Kumar Saha, Rituparna Mitra and Titas Kumar Nag

Abstract

Transmission lines consume a considerable amount of power. The necessity of power and its dependency has grown exponentially over the years. The void between limited production and tremendous demand has increased the focus on minimizing power losses. The losses like transmission loss range from the conjecture factors like physical or environmental losses to severe technical losses. The primary factors like reactive power and voltage deviation are significant in stretched conditions and long-range transmission lines of powers. The subject of power transmission has always been an interest of researchers. A numerous methods have been developed to maximize the throughput of the systems. This paper reviews about different techniques for transmission fault detection.

1. Introduction

Since transmission lines are necessary to transfer power from generating stations to load centres, they are the most significant part of a power distribution system. Transmission lines function at certain voltage levels between 69 and 765 kV and are tightly linked for dependable operation. A multitude of factors, such as the deregulated market environment, right of way, economics, environmental standards, and clearance requirements, force utilities to operate transmission lines near to operating limits. Faults must be found in order to prevent system disruptions, which would then result in widespread outages in the tightly networked system operating within its bounds. “The design of transmission protection systems is in such a way so as to locate the fault location and segregate only the faulted part. It is a very challenging task to identify and isolate the faults in order to have a very reliable transmission line protection.”

The installation of series capacitors becomes highly beneficial for the technical and economic reasons of improved voltage control, supple power-

flow regulation, reduced transmission losses, improved voltage drop, enhanced transmittable power, and increased power system stability [1, 2, 3]. Capacitors arranged in series are used to compensate for the inductive impedance of the line. By utilising a series capacitor, an even better voltage profile can be achieved. Series capacitors lower line inductance by inducing voltage (L di/dt) along the line [4]. The voltage imposed by the source is superposed by the induced voltage. When the line current lags behind the line voltage, there is a voltage drop, and the voltage is increased if the line current leads the voltage. From this angle, series capacitors lower voltage boost and voltage drop because series compensated lines have lower line inductances than uncompensated lines.

A three-phase system experiences a fault that can be either symmetrical or unsymmetrical when two or more conductors come into touch with one another or with ground [5]. Excessive currents resulting from faults cause stress in the equipment of the power system. Moreover, these flaws seriously harm the components of the power system. Faults not only negatively impact the equipment but also degrade the quality of the power. Thus, it becomes essential to determine the type of fault and its location on the gearbox line so that it may be removed with the appropriate means, protecting the power system equipment from harm and improving the quality of the power.

It has proven to be an extremely challenging task to identify and categorise the transmission line issues. Finding and categorising defects is the electricity industry's top priority. To identify the defect and isolate the problematic area from the rest of the system, protective relays, specialised control devices, protection software, and recording devices are essentially employed. To identify the error and fix it as quickly as possible, it is crucial to have complete knowledge about it. Many studies are currently being conducted to learn more about fault location techniques in distribution and transmission networks that are based on artificial intelligence techniques such artificial neural networks and fuzzy-set theory, among others.

2. Faults in transmission lines

Transmission lines are used to distribute power from generating units to distant load centres. Faults on these lines may arise from lightning, malfunctioning, overloading, short circuits, human mistake, defective equipment, and ageing. When a defect arises, the phase voltage drops and massive currents flow, which, if not stopped right away, might burn out the components.

Transmission lines that use a three-phase power supply are susceptible to 10 different kinds of defects. The faults include single-phase-to-ground (L-G), phase-to-phase (L-L), double-phase-to-ground (L-L-G), and a three-phase fault (3-), listed in decreasing order of frequency of occurrence.

Faults that are single line to ground (L-G) happen when a phase is cut short to the ground. Zfag, the impedance, may not be zero (bolted) during the fault, but it will still be substantially lower than the line impedance. While the voltage does not dramatically alter in magnitude, the amount of current in a defective line rises significantly higher than the typical operating current.

Table 1: Types of faults

Types of faults	Symbol	% Occurrence	Severity
Line to ground	L-G	75-80%	Very less severe
Line to line	L-L	10-15%	Less severe
Double line to ground	L-L-G	5-10%	Severe
Three phase	3-ϕ	2-5%	Very severe

Fault location techniques

- Technique based on fundamental-frequency voltages and currents, primarily on impedance measurement
- Technique based on high-frequency components of currents and voltages generated by faults
- Technique based on traveling-wave phenomenon
- Knowledge-based approaches.

The simplest method for pinpointing the fault location is one that takes into account both the line specifications and the fundamental frequency currents and voltages at the line terminal. The common consensus is that the fault's distance can be determined from the impedance calculated for the faulted line section. The methods in this category are easy and inexpensive to use. When performing this kind of categorisation, it is necessary to consider the availability of measurements: whether they come from one or both ends, and if they are used from a certain line end in their entirety or in part (voltage and current).

In traveling-wave methods, the current and voltage waves, traveling at

the speed of light from the fault towards the line terminals are consider [25]. These methods are considered as very accurate, however, also as complex and costly for application, as requiring high sampling frequency."

3. Literature review

It is very important to know the effect of series compensation on transmission voltages [4]. If the effect of series compensation on voltages is not known it will cause various operational problems such as high voltages and low voltages. Series compensation can cause low and high voltages due to different line loading conditions and the method by which the voltage control is adjusted. The voltage on the one side of the capacitor should be adequately controlled otherwise the other end of the capacitor cause voltage problems. When the line is lightly loaded, over-voltages can cause problems and in this case, series compensation will decrease voltage. But in case of heavily loaded lines, low voltages occur across the line, so series compensation will increase the voltage. Voltage collapse results from low voltages. And high voltages either cause flashover or decrease the life cycle of insulation and short circuits. Occasionally, series compensation is used to control power flow."

Current and voltage inversions occur on a series compensated line during directional relaying [1]. Positive-sequence fault voltage and positive-sequence fault current changes in magnitude and phase can be used by the fault identification system for series compensated lines to determine the fault direction. An algorithm is created for a series compensated line using fault current and voltage phasors to acquire the decisions, and simulations are performed using EMTDC/PSCAD. The procedure is tested for various system conditions, fault resistance, changes in source capacity, power-flow direction, fault inception angle, and series compensated lines with and without capacitors.

The distributed parameter double-circuit series-compensated line fault location algorithm was obtained using unsynchronised measurements using the current and voltage data from both ends [6]. Two subroutines are utilised in this approach to locate errors. A variety of formulas are derived using the generalised fault-loop model. The position of the compensating bank is the only factor that affects the fault's distance, not the compensatory bank's specifications. It has been considered to employ two-end signals monitored asynchronously. To simulate various fault types on a double-circuit series compensated line, ATP-EMTP is utilised.

A procedure is presented [7] for figuring out if the detected fault is outside the line or on a protected double-circuit line. Both series-compensated and double circuit uncompensated lines can be protected with it. Phase current measurements are limited to one end measurements. ATP-EMTP software is used for simulation. A comprehensive transmission line model featuring measurement channels, SCs, and MOVs banks is created. The suggested algorithm makes use of two subroutines: one for balancing lines and the other for identifying present imbalance. For symmetrical parallel lines, fault classification is done with near 100% accuracy; for unbalanced lines, it is more than 85-95% accurate.

In [8], a digital distance relaying-based technique for first-zone protection of series compensated double-circuit transmission lines is described. Data from one end of the line is taken into consideration to determine the fault distance. The suggested approach is not dependent on fault current or source impedance. MATLAB/SIMULINK software is used to model a double circuit series compensated 400kV, 300km transmission line. The simulation results demonstrate that the suggested method may be utilised to accurately estimate fault distance.

A suggested algorithm [9] for fault location on series-compensated lines is applicable to both single- and double-circuit series-compensated lines. Current differential relays are taken into consideration at both ends of the line to find a more precise fault locating solution. Differential protection relays can be used in conjunction with fault locators to implement the suggested fault location technique. Differential relays communication infrastructure is used in this way. Therefore, additional channels of communication are not needed. Additionally, the utility of differential relays is greatly boosted.

A suggested algorithm [9] for fault location on series-compensated lines is applicable to both single- and double-circuit series-compensated lines. Current differential relays are taken into consideration at both ends of the line to find a more precise fault locating solution. Differential protection relays can be used in conjunction with fault locators to implement the suggested fault location technique. Differential relays communication infrastructure is used in this way. Therefore, additional channels of communication are not needed. Additionally, the utility of differential relays is greatly boosted.

A novel approach to resolving protection issues with parallel transmission lines is put forth [11]. It is predicated on the three-phase line

voltages and six-phase line currents of the two parallel circuit lines at both ends. Wavelet transform is used for fault analysis. Moreover, internal faults on double circuit lines are identified through a comparison of the magnitudes of the associated current phasors on each line. It is demonstrated that each form of problem may be correctly identified under various loading scenarios.

Accurate fault detection and classification is required in order to distribute power to various places efficiently [12]. The precise protection of transmission lines is ensured by active tripping of circuit breakers, whose tripping action is contingent upon the waveforms of voltage and current during the fault. When analysing current waveforms during a fault, the Discrete Wavelet Transform (DWT) is employed. Discrete wavelet analysis is evaluated for fault identification and classification on a transmission line network. A fault categorisation based on energy level percentage has been completed.

It is detailed how to secure the series compensated line using a current differential pilot relay (CDPR) by using wavelet transform [3]. MATLAB is used to obtain simulation data, and db4 is used as the mother wavelet for analysis. The process of fault classification involves employing a wavelet-based technique to identify various fault kinds. To distinguish between the various kinds of errors, one can also employ the probability-based Bayesian linear discrimination technique [13]. The wavelets are produced via an adaptive wavelet algorithm (AWA) that use a probability-based approach to Bayesian linear discrimination. It is demonstrated that adaptive wavelets can function as analysis filters in the high-speed protection system's gearbox lines.

Transmission of power from the power plant to the distant load centres is required [14]. Therefore, there is a significant chance that the transmission cables will malfunction. This is where signal processing for digital distance protection comes in. Fault location techniques involve the use of wavelet transformations and the Fourier transform. MATLAB/SIMULINK is used for simulation. The results of the simulation indicate that the wavelet method is a more reliable instrument for locating transmission line issues. Furthermore, it is demonstrated that regardless of the noise levels present, the wavelet transform and Fourier transform approaches can both be utilised to determine the features of disrupt signals.

High-speed EHV transmission lines have been protected using discrete wavelet analysis [15]. A discrete wavelet analysis-based fault detection and

classification technique has been introduced. The type of defect is determined by comparing the various wavelet coefficients of the three phase signals. Additionally, the MATLAB Wavelet toolkit and ATP-EMTP are used for simulation. Such a method is described that is independent of the fault location, fault impedance, and fault inception angle. The method is robust, efficient, and well-suited for protecting EHV transmission lines.

A novel method is suggested [2] for the fault classification and boundary protection of series-compensated transmission lines. To have the proper boundary protection, several frequency bands of the transient fault current wave are identified. Using db4 as the mother wavelet, the two frequency bands of the transient fault current signal are gathered. The spectrum energies of two bands of frequencies are calculated to identify if the problem is internal or external. The average value of each current wave's wavelet coefficient is used to categorise failed phases. For all kinds of failures, the fault current values of three phases are used to generate a basic modal signal. Using db4 as the mother wavelet, the modal signal is analysed. The modal signal's detail 1 and detail 6 coefficients are then determined. The ratio of spectral energy and the average values of the d6 coefficients of the ground current and three phase currents are calculated to determine whether the fault is internal or external. These values are then applied to further classify the fault type.

A novel method for TEED transmission line protection is proposed [16], which makes use of wavelet transforms for precise fault location, identification, and classification. In order to generate D1 detail coefficients, the three phase currents at each terminal are broken down at a single level using Bior2.2 as the mother wavelet. The three ends of the TEED transmission line's detail coefficients are then summed to produce the resultant detail coefficients, which are then compared with threshold values to identify and categorise different kinds of problems in TEED transmission lines.

It is explained [17] how neural networks and fuzzy logic are used to secure double circuit series compensated transmission lines. The reliability of the power system is further enhanced by the use of fuzzy logic and neural networks, which are utilised for precise decision-making and for estimating the actual state of the power system, respectively. This enhances the selectivity of the protection system. The study examines the impact of fault resistance, series compensation, and mutual zero-sequence coupling. The

real state of the power system determines how the mutual coupling of a parallel circuit and a series capacitor impedance affects the accuracy of the relays. In the end, it is demonstrated that a fuzzy logic decision-making system reduces the relay sensitivity to nearly nil when the power supply situation changes.

It is crucial to find transmission line faults using an impedance-based calculating method [18]. Line settings can be adjusted to alter the results. This is demonstrated by analysing the two popular approaches in actual faults. The most popular fault location techniques are contrasted, and the relative benefits and drawbacks of each are discussed. The consumers find it easier to select the most precise method in this way. In the conclusion, it is demonstrated that the two end techniques are more robust than the one end ways since the two end methods are less sensitive to faults than the one end methods.

Through decomposition at multiple levels, the Discrete Wavelet Transform (DWT) is utilised to recover the hidden variables from the fault signals [19]. The Daubechies db6 wavelet is employed in single-level decomposition. In order to identify and characterise the faulty phase in ground faults, a threshold is computed. By obtaining local and remote fault information as well as the transmission line's length, the location of the problem can be identified. The system is thought to have very little fault resistance.

An artificial neural network (ANN)-based double circuit transmission line protection technique has been proposed [20]. There are three steps in this method for classifying and recognising different types of defects. Data from one end of the double circuit transmission line has been utilised to calculate the wavelet coefficients. Primary protection is provided to the transmission line as a whole by using only one end of the data. Adjacent transmission lines orientated both forward and backward are provided with backup protection. This technique improves gearbox line protection's first zone reach setting by up to 99% of the line's length.

A wavelet transform-based fault classification system is put forth [21]. To compute dWIabc, current samples from the three lines are utilised. EMTP software does simulation for a variety of fault parameters, including fault inception angles, fault locations, fault lengths, and fault parameters. It is demonstrated that the wavelet transform's magnitude is useful for setting thresholds that allow for the discrimination of various fault kinds and, ultimately, the classification of faults.

C programming is used to create an algorithm based on the discrete wavelet transform [22]. With MATLAB, a 500 kv, 200 km single line is simulated. It is demonstrated that the percentage error increases quickly with increasing fault resistance. Furthermore, when the circuit's reactance is taken into account while determining the distance to a fault, the percentage error in the distance measurement rises as fault resistance does.

Using Biorthogonal and Haar wavelets, a wavelet transform technique is developed to analyse power system disturbances such as transmission line problems [23]. This paper proposes a wavelet transform based method for detecting transmission line defects. To detect transmission line faults and choose the best wavelet for a certain application, one can use the coefficient of discrete approximation of the dyadic wavelet transform using various wavelets as an index.

4. Conclusion

In each research endeavour, conducting a literature review is crucial since it demonstrates the necessity of the task and the background development. It raises pertinent questions about how to improve the previously completed study and permits unresolved issues to surface, clearly defining all boundaries with regard to the project's progression. A large body of research has been evaluated regarding the identification of gearbox line faults. Compensatory circuits are included to enhance the transmission lines' power quality. It is crucial to identify and diagnose the issue quickly, isolate the problematic area precisely, and boost system reliability in order to restore the power supply on time.

References

1. Jena, P. and Pradhan, A.K., "A Positive-Sequence Directional Relaying Algorithm for Series Compensated Line," Power Delivery, IEEE Transactions on, vol.25, no.4, pp.2288-2298, Oct., 2010.
2. Megahed, A.I., Monem Moussa, A. and Bayoumy, A.E., "Usage of wavelet transform in the protection of series-compensated transmission lines," Power Delivery, IEEE Transactions on, vol. 21, no. 3, pp. 1213-1221, July, 2006.
3. Pandya, V. J. and Kanitkar, S.A., "A novel unit protection scheme for protection of series compensated transmission line using wavelet transform," Power Engineering Conference, 2007.
4. Kincic, S. and Papic, M., "Impact of Series Compensation on the

voltage profile of transmission lines," Power and Energy Society General Meeting (PES), pp.1-5, July, 2013.

5. Shaaban, S.A. and Hiyama, T., "Transmission Line Faults Classification Using Wavelet Transform," 14th International Middle East Power Systems Conference (MEPCON'10), Cairo University, Egypt, pp. 532-537, December, 2010.
6. Izykowski, J., Rosolowski, E., Balcerek, P., Fulczyk, M. and Saha, M.M., "Fault Location on Double-Circuit Series-Compensated Lines Using Two-End Unsynchronized Measurements," Power Delivery, IEEE Transactions, vol. 26, no.4, pp.2072-2080, Oct., 2011.
7. Pierz, P., Balcerek P. and Saha, M. M., "A Method for Internal and External Fault Discrimination for Protection of Series Compensated Double-Circuit L_{ine}" Power Tech, IEEE Grenoble, pp. 1-6, 2013.
8. Kumar, R., Sinha, A. and Choudhary, G.K., "A New Digital Distance Relaying Algorithm for First-Zone Protection for Series-Compensated Double-Circuit Transmission Lines," Advances in Computing and Communications (ICACC), 2013 Third International Conference on, pp.102-106, Aug., 2013.
9. Saha, M.M., Rosolowski, E. and Izykowski, J., "A fault location algorithm for series compensated transmission lines incorporated in current differential protective relays" The International Conference on Advanced Power System Automation and Protection, pp. 706-711, 2011.
10. Sharma, R., Ahmad, A. and Shailendra, K. S., "Protection of Transmission Lines using Discrete Wavelet Transform," International Journal of Innovative Technology and Exploring Engineering (IJITEE), vol. 3, Issue-1, June, 2013.
11. Osman, A.H. and Malik, O.P., "Protection of Parallel Transmission Lines Using Wavelet Transform" IEEE Transactions on power delivery, vol. 19, no. 1, 2004.
12. Ashok, V., Bangarraju, K. G. V. S. and Murthy, V.V.N., "Identification and Classification of Transmission Line Faults Using Wavelet Analysis," ITSI Transactions on Electrical and Electronics Engineering, vol. 1, no. 1, pp. 117122, 2013.
13. Perez, F.E., Orduna, E. and Guidi, G., "Adaptive wavelets applied to fault classification on transmission lines," IET Generation, Transmission and Distribution, vol. 5, no. 7, pp. 694.702, 2011.

14. Roshni, U., Niranjan, V., Prakash, C.D. and Srinivas, R., "Location of Faults In Transmission Line Using Fast Fourier Transform And Discrete Wavelet Transform In Power Systems," Undergraduate Academic Research Journal (UARJ), vol. 1, Issue-1, pp. 84-86, 2012.

15. Solanki, M. and Song, Y.H., "Transient Protection of EHV Transmission Line Using Discrete Wavelet Analysis," Power Engineering Society General Meeting, IEEE, vol. 3, pp. 1868-1873.

16. Samapth, B. and Vijayabhaskar, K., "Wavelet Based Protection Schemes for TEED Transmission Circuits," International Journal of Engineering Research and Applications (IJERA), vol. 2, no. 6, pp.931-940, Nov. - Dec., 2012.

17. Nidhi, B., Kumar, R. and Sinha, A., "Neural Network and Fuzzy Logic Based Protection of Series Compensated Double Circuit Transmission line" International Journal of Engineering Research and Applications (IJERA), vol. 3, no. 1, pp.15451551, 2013.

18. Andrade, L.D. and de Leao, T.P., "Impedance-based fault location analysis for transmission lines," Transmission and Distribution Conference and Exposition (T&D), 2012 IEEE PES, pp. 1-6, May, 2012.

19. Saravanababu, K., Balakrishnan, P. and Sathiyasekar, K., "Transmission Line Faults Detection, Classification, and Location Using Discrete Wavelet Transform," International Conference on Power, Energy and Control (ICPEC), pp. 233-238, 2013.

20. Yadav, A. and Swetapadma, A., "Improved first zone reach setting of artificial neural network-based directional relay for protection of double circuit transmission lines" IET Generation, Transmission and Distribution., vol. 8, Issue3, pp. 373-388, 2014.

21. Youssef, O. A. S., "Fault classification based on wavelet transform," Transmission and Distribution Conference and Exposition IEEE/PES, vol. 1, pp. 531-536, 2001.

22. Reddy, R.B., Kumar, M.V., Suryakalavathi, M. and Prasanth Babu, C., "Fault detection, classification and location on transmission lines using wavelet transform," Electrical Insulation and Dielectric Phenomena, pp.409-411, Oct., 2009.

23. Ravindra Malkar, Vaibhav Magdum, Sunil Kumar Gunda,

"Transmission Line Fault Analysis Using Wavelet Theory", International Journal of Engineering Science and Technology (IJEST), ISSN: 0975-5462, Vol. 4, No. 06, June 2012.

24. Mohammad, A.B., "Travelling Waves for Finding the Fault Location in Transmission Lines," Journal Electrical and Electronic Engineering, vol. 1, no. 1, pp.1-19, 2013.
25. R.P. Hasabe and A.P. Vaidya, Detection Classification and Location of Faults on 220 Kv Transmission Line Using Wavelet Transform and Neural Network. International Journal of Electrical Engineering & Technology, 5(7), 2014, pp. 32-44.
26. Vinay Kumar Tripathi and Anil Kumar Bharadwaj, Study of Power Flow and Transmission Capacity In Multi Phase Transmission Lines. International Journal of Electrical Engineering & Technology, 5(10), 2014, pp. 43.55.
27. Izykowski, J., "Fault location on power transmission lines," The Technical University of Wroclaw Press, Wroclaw, Poland, 2008.

Chapter - 34

Modelling and Simulation of Three-Phase Voltage Source Inverter Using a Model Predictive Current Control

Author

Sujoy Bhowmik
Electrical Engineering, Swami Vivekananda University,
Kolkata, West Bengal, India

Chapter - 34

Modelling and Simulation of Three-Phase Voltage Source Inverter Using a Model Predictive Current Control

Sujoy Bhowmik

Abstract

In order to create an optimal system for a three-phase inverter that controls the load current, this study focuses on combining a predictive current control with a three-phase voltage source inverter (VSI). A Finite SetModel Predictive Control (FS-MPC) technique with a resistive-inductive load (RL-Load) for a two-level, three-phase VSI is presented in this work. It has been established that in order to lessen the computational effort that results in numerous possibilities. The system is evaluated with a variety of cases. First, the long prediction horizon system performance is tested. Second, the system's dynamic response is examined when the reference's amplitude changes step-wise. Simulations are performed with MATLAB/Simulink to evaluate an FS-MPC's efficacy for the two-level VSI with RL-load.

Keywords: Finite set-model predictive control, voltage source inverter, RL-load, prediction horizon.

1. Introduction

Voltage source converters have been the subject of extensive study over the past few decades across various industrial sectors for numerous applications. Given the rising energy demands and the need for improved power quality and efficiency, control and power conversion using power electronics have become crucial topics today. Predictive control for power electronics, first introduced in the 1980s, offers several benefits for controlling power converters, such as simplicity in understanding and implementation and adaptability to different types of voltage source converters. Although it requires more calculations than traditional control methods, the advancements in fast microprocessors have made it feasible.

This control method, known as finite state model predictive control (FS-MPC), is used in this paper. It has been effectively applied to a broad range of power converter and drive applications, including three-phase inverters and matrix converters, as well as flux and torque control of induction machines. An example of controlling various variables using a single cost function is shown where current control, minimization of switching frequency, and balancing of dc-link voltages in an inverter are achieved simultaneously. In all these applications, switching states are changed at regular time intervals.

This paper combines a three-phase voltage source inverter (VSI) with predictive current control to optimize the system for controlling load current. The effectiveness and robustness of the proposed control method are demonstrated through simulation results. Section two presents the power converter model used in this paper, beginning with the time-continuous model, which is then discretized. Section three develops the control scheme used to manage the power converter model from section two. MATLAB/Simulink modelling and simulation work are discussed in the penultimate section, and the conclusion is provided in the final section.

This control method is referred to as finite state model predictive control (FS-MPC) because the possible control actions (switching states) are finite, and this strategy is utilized in this paper. FS-MPC has been effectively applied to various power converter and drive applications, including a three-phase inverter, a matrix converter, and flux and torque control of an induction machine. An example of controlling different variables using a single cost function is provided, where the current is regulated while minimizing the switching frequency and balancing the dc-link voltages in an inverter. In these studies, the switching states are altered at regular time intervals.

In this paper, a combination of a three-phase VSI and predictive current control is used to optimize a system for a three-phase inverter that manages the load current. The effectiveness and robustness of the proposed control method are assessed through simulation results.

Section two outlines the power converter model used in this paper, starting with the continuous-time model and then discretizing it. The control scheme developed in section three is used to manage the power converter

model from section two. Matlab/Simulink modelling and simulation work are discussed in the penultimate section. The conclusion is presented in the final section of the paper.

2. Power converter model

a) Voltage source inverter layout

The simplest multiple levels VSI power converter is a two-level VSI three-phase power converter, as it only displays two voltage levels. It is the subject of a clear analysis using RL-Load for predictive control approach. Fig. 1 shows the topology of the inverter taken into consideration in this paper. To operate the circuit, flip S1, S2, S3, S4, S5, and S6. As depicted in Fig. 1, the inverter employs two sets of complementary controlled switches in each inverter phase or leg, S1, S2, S3, S4, and S5, S6. Taking into consideration that, in order to prevent short circuiting, the DC supply, the two switches in each inverter phase or leg function as a complimentary pair.

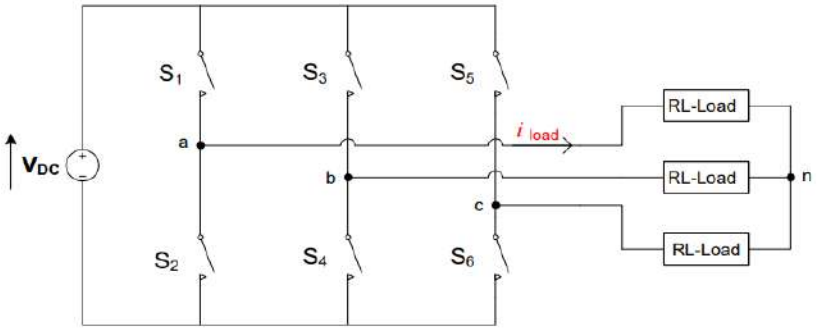


Figure 1: Two level voltage source inverter topology

Eight distinct switching scenarios result from this, and as a result, eight voltage vectors voltages VDC and –VDC that can be switched are produced. A three-phase two-level voltage source converter can only deliver seven different voltage vectors, despite the fact that there are eight different switching combinations, as shown in Fig. 2, because V0 and V7 produce the same zero voltage vector (V0 = V7). This is because the voltage vectors generated by the inverter, as shown in Fig. 2, resulted in only seven different voltage vectors. The following criteria are used to determine the switching states that specify the output voltages:

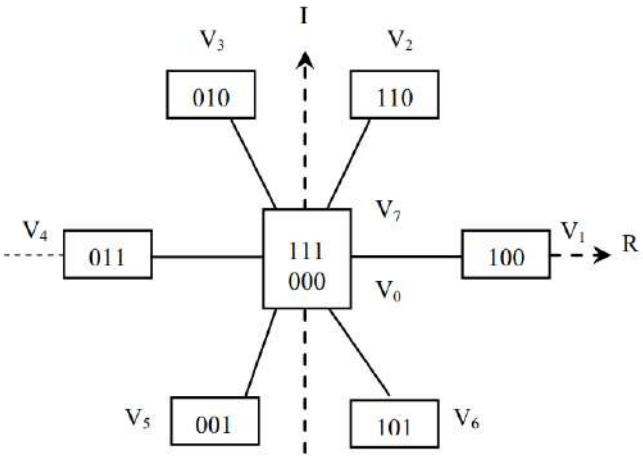


Figure 2: Voltage vectors generated by 2-level VSI

b) Load model

To derive the continuous-time state-space equations of the load for each phase, one applies the differential equation of the load current:

$$Vdc(t) = Ri(t) + L\, di/dt \quad (1)$$

Using the Clarke transformation, the equations can be expressed in the stationary α-β frame. Clarke transformation is defined as following:

$$V_\alpha = 2/3\ (Va - 0.5\ Vb - 0.5\ Vc) \quad (2)$$

$$V_\beta = 2/3\ (0.5\ \sqrt{3}\ Vb - 0.5\ \sqrt{3}\ Vc) \quad (3)$$

A discrete-time equation for the future load current is obtained by using Euler-forward equation as in (4) in order to obtain a discrete-time system representation. The derivation of the state variables is approximated as follows:

$$x^{\cdot} \approx [(k+1) - (k)]/Ts \quad (4)$$

Ts is the sampling time and k is the current sampling instant. The state variable is denoted with x.

To determine the future value of the load current, the cost function g is assessed for each of the seven potential voltage vectors that this inverter may produce. During the subsequent sampling moment, the voltage vector that minimises the cost function is chosen and applied.

3. Control strategy

a) Finite set-model predictive control

A modulator is not required when using an FS-MPC. Current controllers for two- [2-4], three- [13], and four- [14, 15] level inverters have been implemented using FS-MPC [11, 12]. VSI converter under FS-MPCC is depicted in Fig. 3 where: The predictive values of the m states for n possible switching states at time k+1 are represented by i(k+1) and iref, respectively, while m observations at time k are denoted by i(k). The switching state that minimises the cost function is selected by calculating the error between the reference and forecasted values. The converter is then given the switching signals, S, of the selected state.

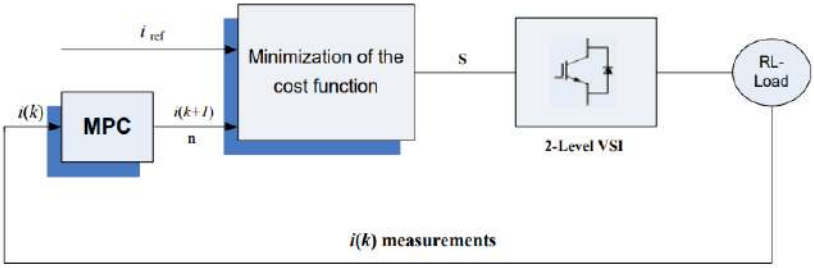


Figure 3: FS-MPC block diagram

The best of the seven voltage vectors is chosen, and the optimal switching state that produces this voltage vector is then chosen in order to minimise the computing work that results in eight distinct switching options for a single prediction step. The 1-norm cost function must be minimised in order to identify which of the seven voltage vectors should be used in the subsequent sampling cycle, k+1. For the current control of a 2-level VSI, a simple cost function with one prediction step and n prediction steps, respectively, can be defined in absolute value terms as follows:

$$g = |\mathrm{i}ref(k+1) - \mathrm{i}\alpha(k+1)| + |\mathrm{i}ref\beta(k+1) - \mathrm{i}\beta(k+1)| \quad (5)$$

$$g = \sum N\ |\mathrm{i}ref(k+n) - \mathrm{i}\alpha(k+n)| + |\mathrm{i}ref\beta(k+n) - \mathrm{i}\beta(k+n)| \quad (6)$$

To streamline the computations, it is possible to consider that the current reference values in equation (7) remain constant, and the prediction horizon for small sampling times TS is represented by equation (8). This work takes this approximation into consideration.

$$\mathrm{i}ref(k+n) \approx \mathrm{i}ref(k) \quad (7)$$

The reference current values $i\vec{re}(k+1)$ for the α-β axis required by (5) must be projected from the current and prior values of the current reference using a second-order extrapolation provided by in order to get a more accurate approximation.

$$iref(k+1) = 3iref(k) - 3iref(k-1) + iref(k-2) \quad (8)$$

Equation (8) has been obtained from the quadratic Lagrange extrapolation formula [3, 16].

In general, the control algorithm as shown in Fig. 3 can be summarized to the following steps [17]:

1) Measure the load currents.
2) Predict the load currents for the next sampling instant for all the possible switching states.
3) Evaluate the cost function for each prediction.
4) Optimal switching state is selected which minimizes the cost function.
5) Apply the new switching state.

In the current control case, the cost function is minimized which obtained from the error between the reference current and predicted current for produce the switching state and applied.

b) Prediction step

As shown in Fig. 3, the controller can use any available output to move the regulated currents closer to their reference by using i(k+1). The controller thus computes the observed currents at k+2 for all conceivable switch states at k+1 in the following step, i(k+2), the prediction step.

There are several options available when an inverter is operated directly. In this instance, the inverter results in eight distinct switching options, which in turn yields eight voltage vectors for a single prediction step. As the prediction horizon lengthens, the computational effort increases to 64 possibilities for two prediction steps and 512 possibilities for three prediction steps. Using MATLAB/Simulink, the potential of a VSI for RL-Load has been assessed with a long prediction horizon. For a three-phase, two-level VSI, Fig. 4 displays the total number of possible switching states with lengthy prediction steps.

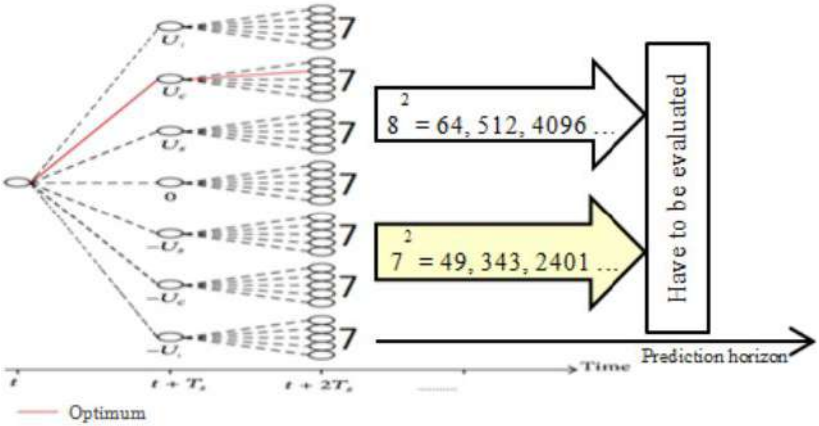


Figure 4: Number of possible switching states with long prediction steps

4. Simulation results

Using MATLAB/Simulink, the FS-MPC strategy for a three-phase, two-level VSI has been simulated in order to assess the effectiveness of the suggested control algorithm and validate the reliability and performance of the suggested predictive control technique. An RL-Load is connected to the output of the VSI as illustrated in Figure 4 after a sinusoidal reference current with an amplitude of 4 A and a frequency of 50 Hz per phase was introduced into the system. Table I. Shows the parameters used for the simulations.

Table 1: Parameters used for the simulations

Parameter	Value
Load resistor, R	10 Ω
Load inductor, L	10 mH
AC Filter, L	10 mH
DC link voltage, VDC	100 V
Reference amplitude current, iref	4 A
Sampling time, Ts	25µs, 75µs

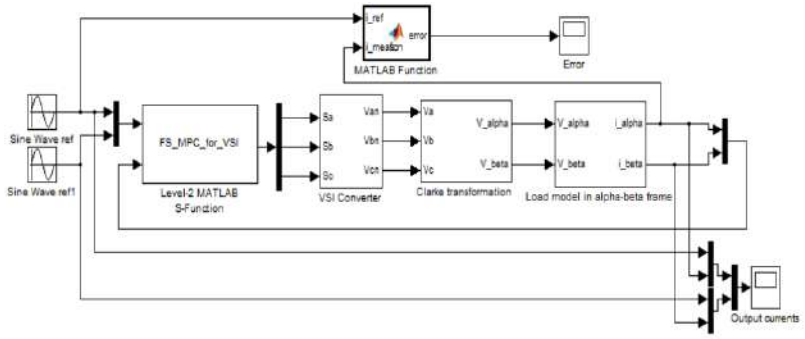


Figure 4: Modelling an FS-MPC of VSI using MATLAB /Simulink

a) With long prediction horizon

Four prediction steps were used to verify the robustness of the suggested control approach. The output currents in Fig. 5 are shown tracking their references with four prediction steps, demonstrating the control algorithm's excellent tracking behaviour.

As illustrated in Fig. 6, the influence of the prediction horizon has been assessed by evaluating the error between the reference and actual current for various values of the forecast horizon.

b) With step change

Using sinusoidal reference steps: The control outcome for sinusoidal reference values is displayed in the second simulation. Figure 7(a) displays the results of the step changes in the amplitude of the references iα and iβ, which were changed from 4 A to 2 A and 4 A to 3 A at 0.042 s and 0.048 s, respectively. Figure 7(b) for the same test shows that the load voltage vα remains at its maximum value throughout the step change of the currents iα and iβ until the reference current iα is achieved. The capacity of the suggested control algorithm to track sinusoidal reference currents is amply demonstrated by this simulation.

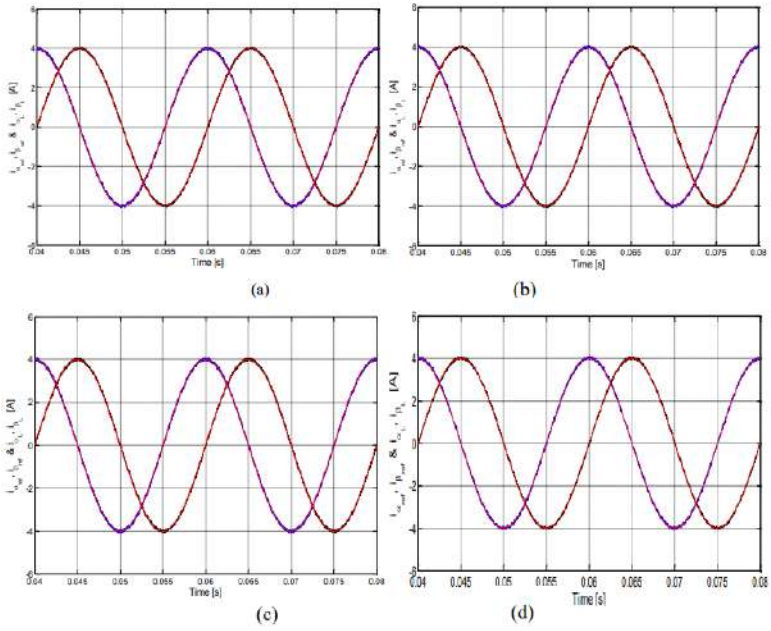


Figure 5: Simulations results: the load current with RL-Load for different values of prediction horizon for a sampling time TS = 25 µs: a) n = 1. b) n = 2. c) n = 3. d) n = 4.

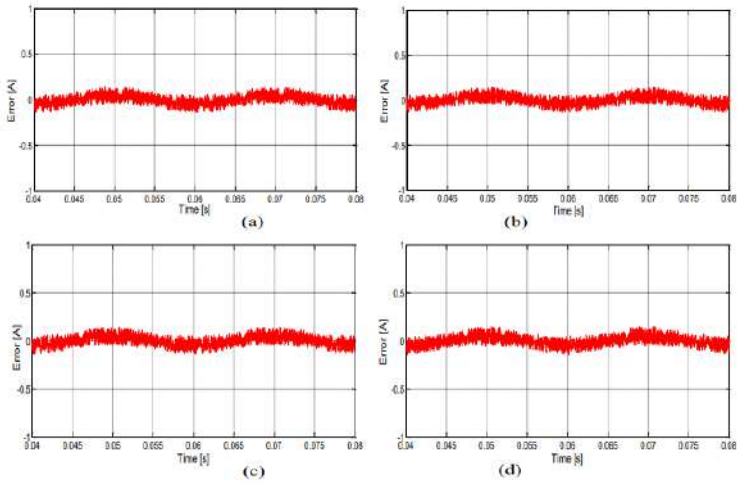


Figure 6: Simulations results: Error between the reference and load current for different values of prediction horizon for a sampling time TS = 25µs: a) n = 1. b) n = 2. c) n = 3. d) n = 4

This simulation shows that the predictive control approach has an intrinsic decoupling between iα and iβ and a quick dynamic response. Once more, the program demonstrates superb tracking behaviour.

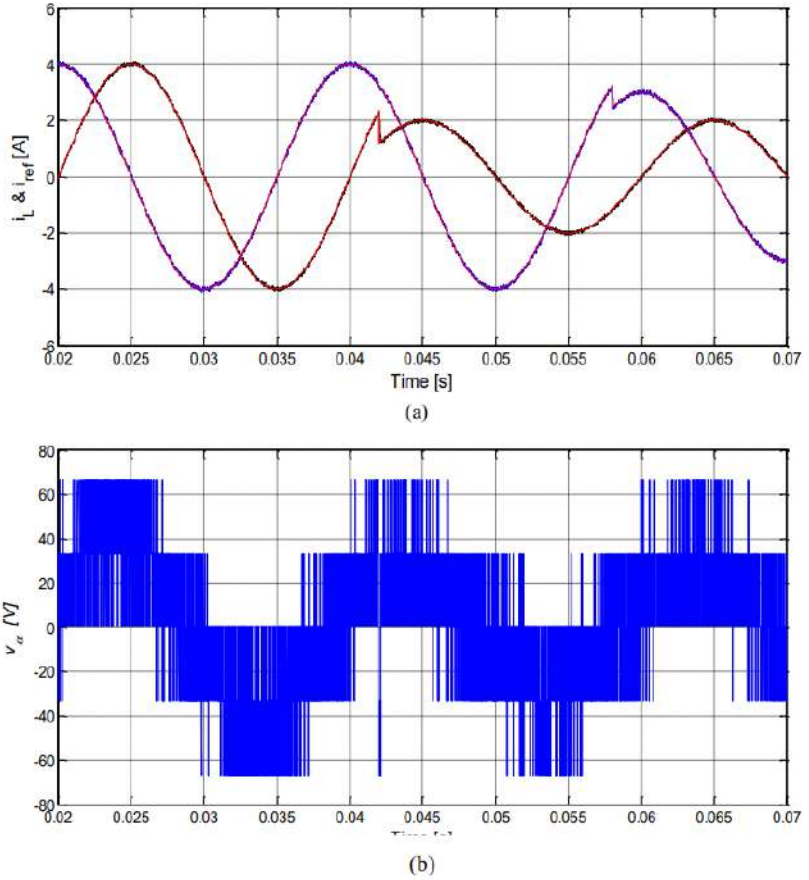


Figure 7: Simulations results: Sinusoidal reference steps: (a) Load current with two step changes. (b) Load voltage vα changes during the reference steps

c) With constant reference steps

In this simulation, constant reference steps were used to test the control algorithm. As seen in Fig. 8, the simulation displays the control result for constant reference steps values when the reference iα and iβ's amplitudes were altered step-by-step from 0 A to 4 A and 0 A to -3 A, respectively, at 0.02 s. Figure 8 illustrates how the load currents quickly and dynamically

approach their reference. The ability of the suggested control method to follow the currents to their reference values is amply demonstrated by this simulation. The finite switching and controller frequency in stable state cause the residual current ripple.

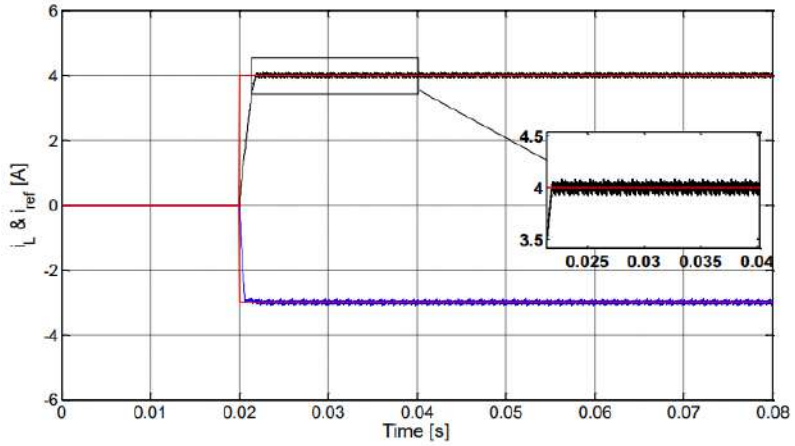


Figure 8: Simulations results: Constant reference steps for load current

5. Conclusions

The FS-MPC for two-level voltage source inverters has been introduced in this paper. The modulator is not required for the suggested control. Through simulation results, three different instances have been used to evaluate the control algorithm. First, the load current and the error between the reference and real current for each of the four prediction steps were checked in order to evaluate the robustness of the proposed control system with four prediction steps. It has been observed that the control algorithm offers excellent behaviour for tracking current. Second, the simulation findings demonstrate that the predictive control approach has a quick dynamic response with intrinsic decoupling between iα and iβ when the reference's amplitude changes step-wise. According to simulation data, the FS-MPC method performs exceptionally well in these circumstances.

References

1. R. Kennel and D. Schroeder, Predictive control strategy for converters, in Proc. the third IFAC Symposium, pp. 415-422, Lausanne, 1983.
2. J. Rodrgí uez and C. Cortés, Predictive control of Power Converters and

Electrical Drives, UK: A Johan Wiley & Sons, Ltd., 2012.

3. J. Rodrgí uez, J. Pontt, C. A. Silva, P. Correa, P. Lezana, P. Cortes, and U. Ammann, ―Predictive current control of a voltage source inverter,‖ IEEE Transactions on Industrial Electronics, vol. 54, no. 1, February 2007, pp. 495-503.
4. H. Abu-Rub, J. Guzinski, Z. Krzeminski, and H. A. Toliyat, Predictive current control of voltage source inverters,‖ IEEE Transactions on Industrial Electronics, vol. 51, no. 3, Jun. 2004, pp. 585-593.
5. J. Li, F. Zhuo, X. Wang, L. Wang, and S. Ni, ―A grid-connected PV system with power quality improvement based on boost + dual-level fourleg inverter,‖ in Proc. IEEE Int. Power Electron and Motion Control Conf., Wuhan, China, May 2009, pp. 436-440.
6. V. Yaramasu, M. Rivera, B. Wu, and J. Rodriguez, ―Model Predictive Current Control of Two-Level Four-Leg Inverters—Part I: Concept, Algorithm, and Simulation Analysis,‖ IEEE Transactions on Industrial Electronics, vol. 28, no. 7, July 2013, pp. 3459 - 3468.
7. R. Vargas *et al.*, ―Predictive control of a three-phase neutral point clamped inverter,‖ IEEE Trans. Ind. Electron., vol. 54, no. 5, pp. 2697-2705, Oct. 2007.
8. S. Muller, U. Ammann, and S. Rees, ―New time-discrete modulation scheme for matrix converters,‖ IEEE Trans. Ind. Electron., vol. 52, no. 6, Dec. 2005, pp. 1607-1615.
9. J. Rodrgí uez, J. Pontt, C. Silva, P. Cortés, S. Rees, and U. Ammann, Predictive direct torque control of an induction machine,‖ presented at Power Electron. Motion Control Conf., Sep. 2-4, 2004, Riga, Latvia,
10. R. Vargas, U. Ammann, J. Rodrgí uez, and J. Pontt, ―Reduction of switching losses and increase in efficiency of power converters using predictive control,‖ in Proc. IEEE PESC, Jun. 15-19, 2008, pp. 1062-1068.
11. E. Camacho and C. Bordons, Model Predictive Control, NY: Springer-Verlag, 1999.
12. P. Cortes *et al.*, ―Predictive control in power electronics and drives,‖ IEEE Transactions on Industrial Electronics, vol. 55, no. 12, Dec. 2008, pp. 4312-4324.

13. P. Stolze, D. du Toit, M. Tomlinson, R. Kennel, and T. Mouton, ―Model Predictive Control of a Flying Capacitor Converter with Output LC Filter for UPS Applications,‖ IEEE Africon 2011 - The Falls Resort and Conference Centre, Sep. 2011.
14. G. Perantzakis, F. Xepapas, and S. Manias, ―Efficient predictive current control technique for multilevel voltage source inverters, ‖ presented at European Conference on Power Electronics and Applications, Sep. 2005.
15. F. Defay, A. M. Llor, and M. Fadel, ―Direct Control Strategy for a Four-Level Three-Phase Flying-Capacitor Inverter,‖ IEEE Transactions on Industrial Electronics, vol. 57, no. 7, 2010, pp. 2240-2248.
16. O. Kukrer, ―Discrete-time current control of voltage-fed three-phase PWM inverters,‖ IEEE Trans. Ind. Electron., vol. 11, no. 2, Mar. 1996, pp. 260-269.
17. P. Cortes, J. Rodriguez, C. Silva, and A. Flores, ―Delay compensation in model predictive current control of a three-phase inverter,‖ IEEE Transactions on Industrial Electronics, vol 59, no. 2, 2012, pp. 1323-1325.