ARTIFICIAL INTELLIGENCE AND DATA SCIENCE IN RECOMMENDATION SYSTEM:

CURRENT TRENDS, TECHNOLOGIES AND APPLICATIONS

Editors:
Abhishek Majumder
Joy Lal Sarkar
Arindam Majumder

Bentham Books

Artificial Intelligence and Data Science in Recommendation System: Current Trends, Technologies and Applications

Edited By

Abhishek Majumder
Tripura University, Tripura, India

Joy Lal Sarkar
Tripura University, Tripura, India

&

Arindam Majumder
NIT Agartala, Tripura 799046, India

Artificial Intelligence and Data Science in Recommendation System: Current Trends, Technologies and Applications

Editors: Abhishek Majumder, Joy Lal Sarkar and Arindam Majumder

ISBN (Online): 978-981-5136-74-6

ISBN (Print): 978-981-5136-75-3

ISBN (Paperback): 978-981-5136-76-0

First published in 2023.

need for a court order if at any point you breach any terms of this License Agreement. In no event will any delay or failure by Bentham Science Publishers in enforcing your compliance with this License Agreement constitute a waiver of any of its rights.

3. You acknowledge that you have read this License Agreement, and agree to be bound by its terms and conditions. To the extent that any other terms and conditions presented on any website of Bentham Science Publishers conflict with, or are inconsistent with, the terms and conditions set out in this License Agreement, you acknowledge that the terms and conditions set out in this License Agreement shall prevail.

Bentham Science Publishers Pte. Ltd.
80 Robinson Road #02-00
Singapore 068898
Singapore
Email: subscriptions@benthamscience.net

CONTENTS

FOREWORD

I have the pleasant task of writing the foreword for the book ***Artificial Intelligence and Data Science in Recommendation System: Current Trends, Technologies, and Applications.*** This is a work edited by Abhishek Majumder, Joy Lal Sarkar of Tripura University, India, and Arindam Majumder of NIT Agartala, India. This book spans certain very crucial and current issues on the theory and application of Artificial Intelligence and Machine Learning. One of the most widely used applications is recommendation systems, which millions of people use on an everyday basis for shopping and entertainment.

The methods in AI and NLP have been in development for several decades. Classification methods and neural networks have also existed for a long time. However, the advent of large-scale gathering of social and user data has recently allowed theoretical techniques to be tested and proved in everyday practice. As a student of AI in the late 80s at IISc, it was difficult for me to imagine this day. We have seen the progression of the methods of pattern recognition and statistical classification methods. There was an interesting twist in the developments of AI systems where in the late 60s, it appeared that linear classification systems and Perceptron training algorithms would progress far. But the failure to solve XOR logic problems led the researchers to believe that these would be ineffective. This has now been very much established to be a fallacy. But the twist took AI research into the development of logic and systems called expert systems. It was imagined that these expert systems would have the real world and the real world experts' knowledge. The knowledge acquisition bottleneck and lack of trainability of the expert systems were their downfalls. There is now a resurgence of another type of system that is filling in this role: the recommender system. These systems are bringing together diverse methods and techniques in AI, Data Science, and large data sets into human interfaces.

Thus, it gives me immense pleasure to see that this compilation has various applications, such as industry 4.0. Going further, we have applications presented here on deep learning, developing applications, movie recommendations, and movie reviews. One of the major applications these days is through natural language processing methods to perform sentiment analysis with data from social media. This is applied to movie reviews for tourist reviews, assessment of prediction and fashion retail, and exploring human resource recruitment and selection aspects. In addition to the very current topics that have been compiled, it is seen that there is a good diversity in the contributors to this volume.

I wish this compilation the best wishes and that the readers might benefit most from it.

Atul Negi
Professor
School of Computer and Information Sciences
University of Hyderabad
Hyderabad
India

PREFACE

A recommendation System is an intelligent computer-based system that serves as a guide and suggests, as per the preferences of the person. It uses state-of-the-art technologies like Big Data, Machine Learning, Artificial Intelligence, *etc*., and benefits both the consumer and the merchant. Recommendation System is becoming very popular as it serves as a guide for the activity that a person or a group plans to perform in the best possible manner, given the constraints imposed by the user(s). Software tools and techniques provide advice on items to be used by a user. The recommendations are to inspire its users to buy different products. This music creation initiative includes specialists in several fields, including Artificial Intelligence, Human-Computer Interaction, Data Mining, Analytics, Adaptive User Interfaces, and Decision Support Systems, *etc*. In this book, the major concepts of recommender systems, theories, methodologies, challenges and advanced applications of recommenders systems are imposed on this diversity. This book comprises various parts: techniques, applications and assessments of recommendation systems, interactions with these systems, and advanced algorithms. The topic of recommendation systems is highly diverse, since it makes it possible for users to make recommendations using different types of user preferences and user needs data. Collaborative filtering processes, content-based methods, and knowledge-based methods are the most common methods in recommending systems. Such three approaches are the basic foundations of recommendation systems. Specialized methods for different data fields and contexts, such as time, place, and social information, have been developed in recent years. Many developments for specific scenarios have been suggested, and techniques have been adapted to different fields of use.

Abhishek Majumder
Tripura University
Tripura
India

Joy Lal Sarkar
Tripura University
Tripura
India

&

Arindam Majumder
National Institute of Technology Agartala
Tripura 799046, India

List of Contributors

Abdul Wahid	Amity Global Institute, Singapore 238466, Singapore
Abhishek Mishra	Indian Institute of Technology, Bhubaneswar, India
Abhishek Majumder	Mobile Computing Lab, Department of Computer Science and Engineering, Tripura University, Tripura, India
Anukampa Behera	Department of Computer Science & Engineering, ITER, S'O'A (Deemed to be) University, Bhubaneswar, India
Alladi Sureshbabu	Department of CSE, JNTUA College of Engineering, Ananthapur, Andhra Pradesh, India
Anupama Angadi	Department of Information Technology, Anil Neerukonda Institute of Technology & Science, Visakhapatnam, Andhra Pradesh, India
André Nascimento	Department of Computing, Federal Rural University of Pernambuco, Recife, Brazil
Balajee Maram	Department of Computer Science and Engineering, GMR Institute of Technology (Autonomous), Rajam, Andhra Pradesh, India
Bibudhendu Pati	Department of Computer Science, Rama Devi Women's University, Odisha, India
Dillip Rout	Department of Computer Science and Engineering, Centurion University of Technology and Management, Odisha, India
Chhabi Rani Panigrahi	Department of Computer Science, Rama Devi Women's University, Odisha, India
Douglas Véras	Department of Computing, Federal Rural University of Pernambuco, Recife, Brazil
Gustavo Callou	Department of Computing, Federal Rural University of Pernambuco, Recife, Brazil
Goddumarri Surya Narayana	Department of CSE, Vardhaman College of Engineering, Hyderabad, TS, India
Joy Lal Sarkar	Mobile Computing Lab, Department of Computer Science and Engineering, Tripura University, Tripura, India
Khushi Chavan	Department of Computer Engineering, Dwarkadas J. Sanghvi college of Engineering, Mumbai, Maharastra, India
Kotrike Rathnaiah Radhika	Department of Information Science and Engineering, BMS College of Engineering, Bangalore, India
Padmaja Poosapati	Department of Information Technology, Anil Neerukonda Institute of Technology & Science, Visakhapatnam, Andhra Pradesh, India
Pundru Chandra Shaker Reddy	Department of CSE, CMR College of Engineering & Technology, Hyderabad, TS, India
Pooja Selvarajan	Department of Computer Science and Engineering, Sona College of Technology, India

Poovizhi Selvan	Department of Computer Science and Engineering, Sona College of Technology, India
Pradeep Kumar	Department of CS&IT, Maulana Azad National Urdu University, Hyderabad, India
Ramchandra Mangrulkar	Department of Computer Engineering, Dwarkadas J. Sanghvi college of Engineering, Mumbai, Maharastra, India
Rajesh Bhatia	Department of Computer Science & Engineering, Punjab Engineering College, Chandigarh, India
Satya Keerthi Gorripati	Computer Science and Engineering, Gayatri Vidya Parishad College of Engineering (Autonomous), Visakhapatnam, Andhra Pradesh, India
Sumi Kizhakke Valiyatra	Institute of Management in Kerala, University of Kerala, Kerala 695034, India
Suneetha Merugula	Department of Information Technology, GMR Institute of Technology, Rajam, Andhra Pradesh, India
Sandeep Harit	Department of Computer Science & Engineering, Punjab Engineering College, Chandigarh, India
Sathiyabhama Balasubramaniam	Department of Computer Science and Engineering, Sona College of Technology, Tamilnadu, India
Santhosh Kumar Balan	Department of Computer Science & Engineering, Guru Nanak Institute of Technology, Hyderabad, Telangana
Shilpa Verma	Department of Computer Science & Engineering, Punjab Engineering College, Chandigarh, India
Shreya Roy	Mobile Computing Lab, Department of Computer Science and Engineering, Tripura University, Tripura, India
Sumit Mitra	Managing Partner, Citizen, Odisha, India
Samudrala Venkatesiah Sheela	Department of Information Science and Engineering, BMS College of Engineering, Bangalore
Tushar Deshpande	Department of Computer Engineering, Dwarkadas J. Sanghvi college of Engineering, Mumbai, Maharastra, India
Vidhushavarshini Sureshkumar	Department of Computer Science and Engineering, Sona College of Technology, India
Venkatesh Naganathan	Amity Global Institute, Singapore, Singapore
Yadala Sucharitha	Department of CSE, CMR Institute of Technology, Hyderabad, TS, India

CHAPTER 1

Study of Machine Learning for Recommendation Systems

Tushar Deshpande[1,*], **Khushi Chavan**[1] and **Ramchandra Mangrulkar**[1]

[1] *Department of Computer Engineering, Dwarkadas J. Sanghvi college of Engineering, Mumbai, Maharastra, India*

Abstract: This study provides an overview of recommendation systems and machine learning and their types. It briefly outlines the types of machine learning, such as supervised, unsupervised, semi-supervised learning and reinforcement. It explores how to implement recommendation systems using three types of filtering techniques: collaborative filtering, content-based filtering, and hybrid filtering. The machine learning techniques explained are clustering, co-clustering, and matrix factorization methods, such as Single value decomposition (SVD) and Non-negative matrix factorization (NMF). It also discusses K-nearest neighbors (KNN), K-means clustering, Naive Bayes and Random Forest algorithms. The evaluation of these algorithms is performed on the basis of three metric parameters: F1 measurement, Root mean squared error (RMSE) and Mean absolute error (MAE). For the experimentation, this study uses the BookCrossing dataset and compares analysis based on metric parameters. Finally, it also graphically depicts the metric parameters and shows the best and the worst techniques to incorporate into the recommendation system. This study will assist researchers in understanding the summary of machine learning in recommendation systems.

Keywords: F1-measure, Machine learning, Mean absolute error (MAE), Nearest k- neighbors (KNN), Non-negative matrix factorization (NMF), Recommendation system, Root mean squared error (RMSE), Singular value decomposition (SVD).

INTRODUCTION

Recommendation System

The recommendation system [1] is the main part of digitization as it analyses the interest of users and recommends something based on those interests [2 - 5]. The aim of these systems is to reduce information overload by retrieving the most sim-

* **Corresponding author Tushar Deshpande:** Department of Computer Engineering, Dwarkadas J. Sanghvi college of Engineering, Mumbai, Maharastra, India; Tel: +91-07599029823; E-mail: tushdeshpande791@gmail.com

Abhishek Majumder, Joy Lal Sarkar & Arindam Majumder (Eds.)

ilar items depending on the customer's interest [6 - 10]. The primary use of these systems is decision making, maximizing profits, and reducing risks. This reduces customer's efforts and time in information searching. It works as a filter that suggests alternatives based on massive data. Moreover, it acts as a multiplier that contributes to the expansion of the client's options [11 - 22].

Over the last few years, the enthusiasm for recommendation systems has increased tremendously [23]. This is the most widely used service on high-end websites like Amazon, Google, YouTube, Netflix, IMDb, TripAdvisor, Kindle, *etc.* A number of media companies develop these systems as a service model for their clients. Furthermore, the implementation of such systems at commercial and non-profit sites attracts the attention of the customer [24 - 32]. These also satisfy clients more with online research results. These systems help customers search for their loved items faster and acquire more authentic predictions leading to higher sales at an eCommerce site.

Regarding knowledge of these systems, there are various undergraduate and graduate courses at institutions around the world. Conferences, workshops, and contests are organized in accordance with these systems [33 - 47]. One of the competitions was the Netflix Prize, organized around machine learning and data mining. In this competition, participants were required to develop a movie recommendation system whose accuracy is 10% more precise than the existing system, also known as Cinematch. After a year of hard work, the Korbell team won first place using the two main algorithms: matrix factorization (Singular value decomposition (SVD)) and Restricted Boltzmann machines (RBM).

Real applications [2] employ different ML algorithms, such as K-nearest neighbor (KNN), Naive Bayes, Random Forest, Adaboost, Singular value decomposition (SVD), and many others. The evolution of the recommendation scheme has led to the application of ML and AI algorithms for effective prediction and accuracy. In addition, the results provided by some ML algorithms are expected to be slightly promising. Due to the broad classification of ML algorithms, the choice of an ML algorithm may become a challenge depending on the different situations where recommendation systems are needed. To select an effective ML algorithm, the best way for the researcher or programmer would be to have a thorough knowledge of ML and recommending systems [48, 49]. This knowledge enables the researcher to create a model appropriate to a specific problem. Here, the study provides an overview of ML briefly.

Machine Learning

Machine learning demonstrates the imitation of human learning in computers by learning from experiences and applying them to recently encountered situations.

ML originated in the 1950s but became more popular in the 1990s. Humans understand, but on the other side, the computer uses algorithms.

Machine Learning is classified into four categories:

1. Supervised learning

2. Semi-supervised learning

3. Unsupervised learning

4. Reinforcement learning

Supervised learning

This learning deals with algorithms that provide training data with a set of features and the correct prediction according to those features. The task of the model would be to learn from this data and apply the information learned into new data with the input features and predict its outcome. An example would be predicting the price of a house according to the area.

Semi-supervised learning

In this learning, the model learns from training data that includes missing information. These types of algorithms focus more on concluding from insufficient data. An example is the evaluation of movies where not all viewers will give a review, but the model ends with the reviews provided.

Unsupervised learning

This learning focuses on algorithms that do not require training data. These algorithms use real-world information to learn by themselves. It focuses primarily on relations hidden in the specified data. An example is YouTube, which parses the viewed videos and recommends similar videos to the user.

Reinforcement learning

This type of learning involves algorithms that learn from feedback from an external body. It is similar to a student and teacher where the teacher may give fewer grades (negative feedback) or more grades (positive feedback). An example is to offer a treat to a dog for a positive response and not give that treat for a negative one.

CHAPTER 2

Machine Learning Approaches for Text Mining and Spam E-mail Filtering: Industry 4.0 Perspective

Pradeep Kumar[1,*], **Abdul Wahid**[2] and **Venkatesh Naganathan**[2]

[1] *Department of CS&IT, Maulana Azad National Urdu University, Hyderabad, India*

[2] *Amity Global Institute, Singapore 238466, Singapore*

Abstract: The revolution of Industry 4.0 will leave an impact on the domain of everyone's lives directly or indirectly. Several new complex applications will be developed in the days to come that are complicated to predict in the current scenario. With the help of machine learning approaches and intelligent IoT devices, people will be relieved from extra overheads of redundant work currently being performed. Industry 4.0 has become a significant catalyst for innovation and development in various industrial sectors like production processes and quality improvement with greater flexibility. This chapter applied different machine learning algorithms for spam detection and classifying emails into legitimate and spam. Seven classification models: Decision Trees, Random Forest, Artificial Neural Network, Gradient Boosting Machines, AdaBoost, Naive Bayes, and Support Vector Machines are applied. Three benchmark spam datasets are extracted from standard repositories to conduct the experiments. The chapter also presents a quantitative performance analysis. The results from rigorous experiments reveal that ensemble methods, Gradient Boosting and AdaBoost, outperformed other methods with an overall accuracy of 98.70% and 98.18%, respectively. The ensembled models are effective on a large-sized dataset embedded with more extensive features. The performance of non-ensemble methods, ANN and Naïve Bayes, was instrumental on large datasets as a viable alternative, with an overall accuracy of 98.38% and 97.63% on test data.

Keywords: Cross-validation, Industrial revolution, Machine learning methods, Parameter optimization, Performance measurement, Preprocessing techniques.

INTRODUCTION

With the advancement of Information and Communication Technologies (ICT), Fourth Industrial Revolution (Industry 4.0) embodies several aspects of cutting-edge technology ranging from servicing robots attending to patients during the

* **Corresponding author Pradeep Kumar:** Department of CS&IT, Maulana Azad National Urdu University, Hyderabad, India; E-mail: pradeep@manuu.edu.in

Abhishek Majumder, Joy Lal Sarkar & Arindam Majumder (Eds.)

ever-challenging COVID-19 epidemic situation, Unmanned Aerial Vehicle (UAV), auto-piloted planes, and cars roaming the skies and roads, and entertainment through audio-visuals. Industry 4.0 is a new technological revolution that provides leverage of cyber-physical systems to improve new areas of development based on traditional industrial technology and services through the combination of information and industrialization. Growth and diffusion of Industry 4.0-related technology, such as augmented reality, have provided novel audio-visuals and facilitated several online training programs for professionals like doctors, front-line workers, and volunteers during lockdown periods across the globe. The Internet of Things (IoT) can comprehensively monitor appliances through intelligent sensors for smart cities, highways, and agriculture. Industrial cyber security platforms can perform intelligent monitoring of corporate networks and can take countermeasures against different attacks can be done. The prevalence of the industrial 4.0 revolution promises to standardize and streamline product manufacturing.

Text categorization plays a significant role in text retrieval, information extraction, and question-answering patterns. Intelligent classifiers are found to be more promising for the automatic filtration of text documents. One of the most widely used ways of digital communication is email for personal and business purposes. Therefore, a substantial need is to categorize emails as spam or ham. Spam filtering is a technique to detect unsolicited emails that prevents them from delivering to the user's inbox.

A typical fourth Industrial revolution scenario/platform comprises data and machine learning techniques for a better understanding of the user, product manufacturing, monitoring of the quality of the product, and distribution of logistics with user feedback. Data is captured through the sensors, transferred to the internet's cloud server, and analyzed through machine learning algorithms (supervised, unsupervised, and reinforced). Moreover, intelligent decisions are made for Industry 4.0 users, such as frequency of use, preferences, modes of use, and other related schedules. In addition to statistical data analysis, machine learning and machine vision technology can be applied for automatic large-scale, highly accurate product inspections, particularly identifying complex defects that are not easily visible to the human eye. One of the most diverse field of Industry 4.0 revolution is auto-pilot system. The auto-pilot will play a crucial role in logistic distribution. Within the next decade, it is expected that machine learning, computer vision, and control technology will get fully commercialized for the automated driving technology that will make delivery and logistics much more straightforward while significantly reducing costs. Fig. (**1**). shows the various phases of the industrial revolution.

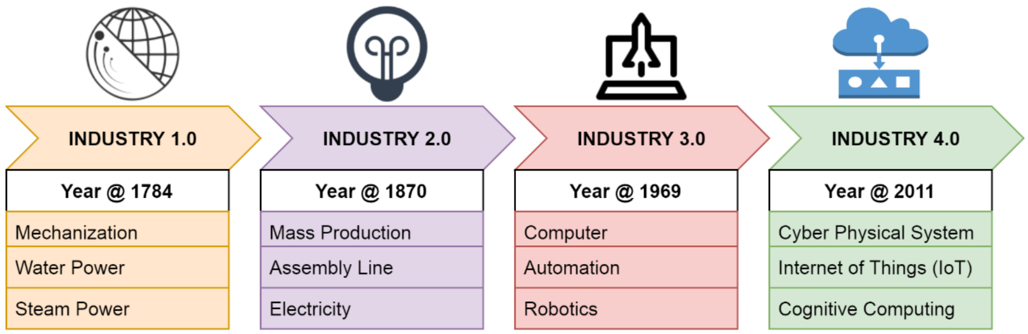

Fig. (1). Industrial revolutions.

Current knowledge about the capabilities needed in Industry 4.0 is inadequate. This research aims to unfold an outline of Industry 4.0 email spam filtering, employing text mining on intimately available email spam filtering, frequently applied as a channel for gathering potential information. The distinguishing characteristics of the fourth industrial revolution primarily include:

Integration and Interconnection

From Industry 4.0 perspective, sensors are integrated and embedded into the hardware. With the help of machine learning engines, everything becomes interconnected. That is, it becomes quite possible to connect people to people (P2P), machines to machines (M2M), people to machines (P2M), and services to services (S2S) seamlessly. Therefore, using integration and interconnection factor, all the processes from production to services like equipment, production lines, factories, and services can be closely linked together efficiently and effectively.

Data and Digitalization

Data components include various aspects of production and associated services like equipment data, product data, supply chain data, operational data, R&D, and user-related data. Machine learning algorithms must be trained and tested with sufficient potential data for efficient outcomes. Deploying machine learning algorithms requires data generation to control the production processes. Thus, using the digitization of data, associated processes can be automated.

Refinement and Personalization

Fourth Industrial Revolution interpolates relatively specific and customized requirements that are increasingly becoming more refined. Each part of the production line becomes more modular and refined, leading to personalized

CHAPTER 3

An Overview of Deep Learning-Based Recommendation Systems and Evaluation Metrics

Samudrala Venkatesiah Sheela[1,*] and **Kotrike Rathnaiah Radhika**[1]

[1] *Department of Information Science and Engineering, BMS College of Engineering, Bangalore, India*

Abstract: The ever-increasing information on the internet and the rapid development of online movies, songs, and stores have enhanced the demands of customers to obtain the required information within the least time. A Recommendation System (RS) is designed to help customers with personalized information and interests to avoid overloading issues in entertainment and social media. Though traditional methods have made noteworthy developments, RS encounters challenges such as data limitations and cold starts. The present study aims to review the developments in the field of deep learning-based RS, thereby providing the required information for researchers. In addition, several applicable domains of employing deep learning-based RS have been analyzed. The review has been organized into RS type, deep learning approaches, deep learning-based recommendation systems in various applications, and evaluation metrics.

Keywords: Content-based approach, Collaborative filtering, Deep Learning, Discriminative approach, Generative approach, Recommendation systems.

INTRODUCTION

Recommendation Systems (RSs) are models that efficiently facilitate personalized information and services to users when data is enormous to explore [1]. Alternatively, RS is defined as a filtering system that overcomes the challenges encountered during an overload of information by selecting the required data based on the behavior and preferences of users [2]. RS is widely used in online shopping, social networking, news sites, research institutions, movies, songs, travel destinations, and e-learning.

RS are classified into three groups based on the recommendations made by users: Content-Based (CB), Collaborative Filtering (CF), and hybrid. Other types incl-

* **Corresponding author Samudrala Venkatesiah Sheela:** Department of Information Science and Engineering, BMS College of Engineering, Bangalore, India; E-mail: pssv.ise@bmsce.ac.in

ude Graph-Based (GB) and Matrix-Based (MB) systems. CB deals with the users' description, features, and profile information; CF utilizes the history of user-based and item-based interactions to produce the required recommendations [3]. GB generates recommendations by finding the ancillary relationships in the graph. The MB methods generate a profile for individual users based on user-item interactions used by researchers for data representation [4, 5]. Singular Vector Decomposition (SVD) is an MB system that analyzes the relationship between users and items and generates a low-dimensional representation of the user-item space. The combined use of CB, CF, GB, and MB approaches to data representation, referred to as the hybrid approach, has been investigated [6, 7].

As shown in Fig. (**1**), the RS framework incorporates data collection, pre-processing, deep-learning technologies, recommendations, and comparison with existing methods for accuracy and performance.

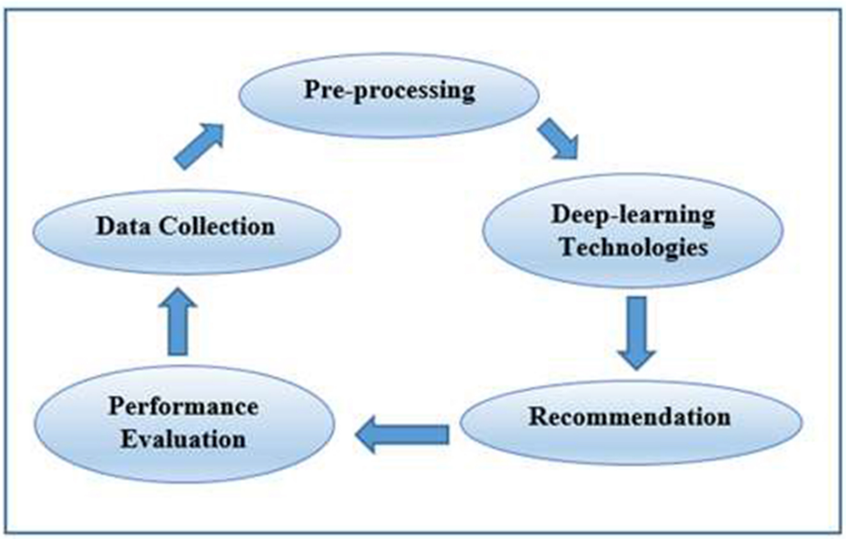

Fig. (1). Recommendation System Framework.

RECOMMENDATION SYSTEMS

Content-based Recommendation

CB recommendation is classified as simple and composite based on the extracted metadata [8]. While articles, photos, blogs, and videos are classified under simple metadata, boards with related groups of messages, questions, and associated answers are grouped under composite metadata. The recommendation is accomplished based on the user profile and item description. The system makes predictions based on the history of user citations instead of users' similarities. Classifications are done regarding item representation, feature learning, and recommendation list generation. Though CB recommendation has a relatively shorter response time and faster speed, lack of enough recommendations and accuracy processing due to extensive data are some of the disadvantages encountered [9].

Challenges faced by CB include the limited association of content with the items, cold-start issues, over-specialization, and higher time consumption. Limited content refers to the scarcity of features to analyze a document, over-specialization refers to similar documents with different titles, and cold-start refers to the shortage of information concerning users and items.

Collaborative Filtering Recommendation

The CF technique, an effective and widely employed recommendation technique, has been introduced to rectify limitations encountered by the CB approach. Unlike CB, CF is implemented based on similarity among users or items [10]. The similarity is calculated using algorithms, such as cosine-based and Pearson correlation [11]. CF builds a user-item matrix based on the priorities/ preferences of users and is classified as memory-based and model-based [12, 13]. Profiles of users/items and user ratings, referred to as memory-based recommendations, are often used as an information store for RS [14]. CF is classified into user-based and item-based depending on whether the similarity is between users or items. For instance, keyword-aware service is a user-based technique, and location-aware RS is item-based [15, 16]. While memory-based RS uses statistical methods, model-based RS uses data mining and machine learning methods. The model-based filtering technique uses previous ratings to learn and apply the CF method. Fuzzy algorithms, genetic algorithms, neural networks, Bayesian networks, and Singular Value Decomposition (SVD) are model-based methods. The neighbor algorithm is a widely used CF method [17]. The reinforcement learning-based RS recommends desired music using user profiles and predicted playlists with higher accuracy than prevailing methods [18].

Despite the significance of CF methods, several limitations are present in systems, including scalability, sparsity, gray-sheep user issue, and cold-start problem [19]. Scalability is a feature of the RS that describes its ability to cope and execute efficiently when faced with an increasing or expanding user number. The system can scale well, maintain or even improve its performance level when more users test it. Data sparsity is based on massive datasets. Consequently, user-item matrices employed for CF may be vast and sparse, bringing on difficulties in performance. Gray sheep users are those whose perceptions differ and may neither

CHAPTER 4

Towards Recommender Systems Integrating Contextual Information from Multiple Domains through Tensor Factorization

Douglas Véras[1,*], **André Nascimento**[1] and **Gustavo Callou**[1]

[1] *Department of Computing, Federal Rural University of Pernambuco, Recife, Brazil*

Abstract: Traditionally, single-domain recommender systems (SDRS) can suggest suitable products for users to alleviate information overload. Nonetheless, cross-domain recommender systems (CDRS) have enhanced SDRS by accomplishing specific objectives, such as improving precision and diversity and solving cold-start and sparsity issues. Rather than considering each domain separately, CDRS uses information gathered from a particular domain (*e.g.,* music) to enhance recommendations for another domain (*e.g.,* films). Context-aware Recommender System (CARS) focuses on optimizing the quality of suggestions, which are more appropriate for users depending on their context. Integrating these techniques is helpful for many cases where knowledge from several sources can be used to enhance recommendations and where relevant contextual information is considered. This work describes the main challenges and solutions of the state-of-the-art in Cross-Domain Context-Aware Recommender Systems (CD-CARS), taking into account the abundance of data on different domains and the systematic adoption of contextual data. CD-CARS have shown efficient methods to tackle realistic recommendation scenarios, preserving the benefits of CDRS (regarding cold-start and sparsity issues) and CARS (assuming accuracy). Therefore, CD-CARS may direct future research to recommender systems that use contextual information from multiple domains in a systematic way.

Keywords: Recommender System, Cross-domain, Context-awareness, Collaborative filtering, Tensor factorization.

INTRODUCTION

Nowadays, the rapid increase of Internet services has been growing the amount of data available, and the activity of discovering meaningful information has become an issue known as the *information overload problem* [1 - 6]. Assuming the enormous amount of information available through Internet services, the definition of *information* is essential. We can define *information* as any *item* that can be

* **Corresponding author Douglas Véras:** Department of Computing, Federal Rural University of Pernambuco, Recife, Brazil; E-mail: douglas.veras@ufrpe.br

Abhishek Majumder, Joy Lal Sarkar & Arindam Majumder (Eds.)

adopted and rated by any customer (products, movies, *etc.*). Therefore, the information overload problem turns this procedure hard for anyone.

Researchers from the Artificial Intelligence subject have been driving their attention to this information overload issue, so recommender systems (RSs) have been proposed [6]. Recommender systems (RS) can suggest videos to watch, books to read, and so on. These recommendations are given considering the users' records, which could be derived, for example, from usage logs.

A rising number of software and websites have been considering recommender systems to give customers relevant items, such as Amazon, Netflix, and Youtube. Indeed, most of these systems are created for suggesting items in specific domains (*e.g.,* movies, videos). These systems are named *single-domain* RS (SDRS) because they assume the user profile from only one domain to produce the suggestions. For instance, an SDRS may suggest films depending on the previous one watched. In this work, a "domain" is assumed as a "type of product" (*e.g.,* music, movies) [7].

Notwithstanding SDRS has obtained significant results for the suggestions provided, issues are still open [6]:

● Cold-start: Situations where RSs are unable to give suggestions due to the absence of initial data or preferences;

● Sparsity: The mean number of ratings is low, which results in low-quality suggestions;

● Diversity: Similar (or duplicated) items can be recommended, which may reduce the satisfaction level of customers;

● Accuracy: non-consistent predictions may be provided, and the recommended items list may not be enough to satisfy users.

To mitigate such issues, *cross-domain* recommender systems (CDRS) [7] have appeared, intending to optimize the quality of suggestions provided by single-domain systems. Rather than dealing with each domain separately, CDRS adopts expertise obtained from a domain (*e.g.,* books) to enhance suggestions for another domain (*e.g.,* movies). For instance, the issues related to the lack of information on favorite movie genres can be reduced by assuming the favorite book genres. *Winoto et al.* [8] analyzed that the consumption behaviors of some items may be effective in providing suggestions for a specific domain. In this research, the authors showed that while recommendations from several domains could not be precise, they can be far more diverse than suggestions from only one domain.

Many other methods have been created since the first cross-domain recommender systems were proposed [7]. For instance, Trewin *et al.* [9] presented a knowledge-based recommender system to give suggestions using the knowledge of products and customers, including their connections. A lot of data knowledge should be available and organized to infer and reason. However, this knowledge extraction is not a simple process, requiring a knowledge engineer who can build the knowledge base and represent a bottleneck for these recommender systems [10].

Fernández-Tobías *et al.* [11] stated that domains could be connected through content-based (CBF) or collaborative filtering (CF) particularities related to customers and products (*e.g.,* preferences, social notes). Cremonesi *et al.* [12] observed and classified *cross-domain collaborative filtering recommender systems* (CD-CFRS) as depicted in Fig. (**1**). CF has been considered the most advanced method in cross-domain RS since its application is relatively simple, and the quality results are usually above that of other approaches [11].

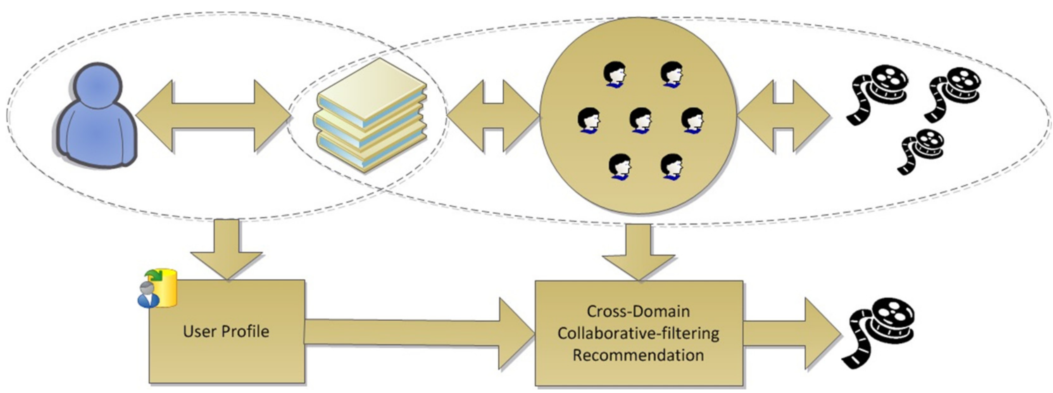

Fig. (1). Cross-domain collaborative filtering recommendation (CD-CFRS).

Indeed, the main part of the CD-CFRS provides higher recommendations than single-domain RS, dealing with cold-start, sparsity, and diversity issues. Besides that, CD-CFRS has higher accuracy than single-domain collaborative filtering RS [11]. Context-aware recommender systems (CARS) represent an important study area of recommender systems [13]. It can provide suggestions of better quality by giving accurate recommendations that consider the user's context.

Fig. (**2**) depicts the context-aware method that adopts distinct contextual data (*e.g.,* location, period, mood) to optimize the suggestions' correctness [13]. For instance, to provide a suggestion for a trip on a vacation or a simple movie, it may be necessary to consider contextual data about the recommendation process in addition to customers and product preferences [13]. A trip recommender system

CHAPTER 5

Developing a Content-based Recommender System for Author Specialization using Topic Modelling and Ranking Framework

Shilpa Verma[1,*], **Rajesh Bhatia**[1] and **Sandeep Harit**[1]

[1] *Department of Computer Science & Engineering, Punjab Engineering College, Chandigarh, India*

Abstract: In the era of information overloading, enormous scholarly data poses a challenge in identifying potential authors for productive outcomes. Researchers collaborate with fellow profiles to improve the eminence of Research and their academic profiles. This chapter proposes a content-based recommender system to generate author recommendations for collaborations that extracts the relevant keywords from the titles of research papers using MapReduce. To specify author specialization, the proposed technique comprises the feature extraction from the entire document using Latent Dirichlet Allocation (LDA), followed by an influence model which generates recommendations for the target authors. A ranking algorithm, such as TOPSIS is implemented to get Top-N recommendations based on multiple criteria. In this chapter, we investigated how the MapReduce framework is helpful in obtaining improved computational time for large-scale scholarly data and scalability. Experimental results on DBLP articles prove the relevance of ranking methods as an efficient and scalable platform for computing content-based recommendations.

Keywords: Author Recommendation, Hadoop MapReduce, PageRank, Cold-Start issue.

INTRODUCTION

With the rapid growth of scientific data in the real world, researchers spend considerable time and effort to get satisfying outcomes as research papers and connections. The contribution of researchers is increasing with time as an academic community publishes millions of articles and connects with multiple researchers. Several vital tasks come in the way of researchers, from finding topics to gathering literature on the relevant area in which one is researching. In this process, researchers rely on the manual process of listing the names of auth-

* **Corresponding author Shilpa Verma:** Department of Computer Science & Engineering, Punjab Engineering College, Chandigarh, India; E-mail: shilpa@pec.edu.in

Abhishek Majumder, Joy Lal Sarkar & Arindam Majumder (Eds.)

ors who share the same research area. It causes researchers a lot of time waste and makes it challenging to manage resources that continue to grow quantitatively. However, studies suggest that more collaborations lead to incredible popularity in the research community. In such a scenario, an author of a scientific article requires a set of refined lists of authors in the relevant area, which will be a valuable asset for the author. There are various motivations for a researcher behind this service of providing collaborators:

a. Searching the literature relevant to the topic
b. Finding the relevant references to be used in the articles related to the author's interest
c. Finding the list of likely reviewers which are required by journals
d. Following the related researcher's work

Finding specialization and relevant authors based on limited constraints is a challenging task. The Hadoop [1] based framework is effective for mining the text. It is instrumental in solving the problem of large-scale scholarly documents. Latent Dirichlet Allocation (LDA) [2] is an unsupervised topic modeling method in which hidden users' preference is reflected by their published articles. However, to reduce workload and improve the computational efficiency of topic modeling, we adopt a topic modeling method with a parallel computing framework. The parallel computing framework of MapReduce can achieve better fault tolerance, load balancing, and scalability. The title in the MapReduce methodology is used to find an author's specialization domain. LDA [3] is used to extract the features of the topic by analyzing the inherent patterns of all documents. It would help new authors with no history to find fellow researchers/experts as a topic profile is created from published articles using topic modeling.

Moreover, considering some of the attributes for recommendation cannot effectively provide the domain expertise of the author. Other attributes related to the author entity, such as references, venue, citation number, abstract, *etc.*, need to be considered to improve the accuracy of specialization. Therefore, many studies used PageRank centrality to understand the correlation and author's importance. Ma et al. [4] compared paper citations with PageRank value. They have shown that PageRank eliminates the effect of self-citations. Thus, in this chapter, using the reference of PageRank centrality with a parallel topic modeling framework, recommendations and ranking of domain experts have been made from a huge scholarly dataset, excluding self-citations.

Additionally, because of incompatible criteria proportions, normalization is often needed in multi-criteria problems. TOPSIS [5] is one of the compensatory methods that tolerates trade-offs between various criteria, where a positive-ideal solution in one criterion can disprove a negative-ideal solution in another criterion. A more accurate form of modeling than non-compensatory methods is provided, which comprise or reject the ranking of alternate solutions based on the criteria considered. TOPSIS is applied to various applications, such as supplier selection problems, selecting and evaluating the best faculty candidate, assessing the status of competing companies based on relative performance, and selecting the most suitable candidate to hire. In summary, the key contributions of the present work are as described below:

a. This approach exploits the scholarly corpus features that effectively evaluate the author's impact and ranking.
b. Evaluates the academic relationship model that will contribute a significant part to recommendations
c. Compared to single feature-based rankings, it validates the usefulness of the accumulated ranking.

The remainder of the chapter is constructed as follows: Section 2 provides related work on content-based recommendation and influence models. Section 3 presents the problem statement for our content-based recommender system. Section 4 explains the topic analysis framework based on Hadoop with LDA scheme, and Section 5 explains influence models in detail. In Section 6, experiments to evaluate the performance of the proposed recommender system are presented. Finally, the Conclusion and future work are discussed in Section 7.

RELATED WORK

Recommender systems [6] have become an important research area because of the richness of real-world applications such as movies, restaurants, Facebook, and Amazon that benefit the user in dealing with information overload issues. It recommends items for users in a particular domain by estimating a utility function that automatically predicts how a user will like an item. Classically, recommender systems are categorized as content-based filtering(CBF) [7] and collaborative filtering(CF) [8]. CBF is a domain-dependent approach that analyses a set of descriptions of items and recommends items that are similar to other items. In content-based recommendation methods, the utility of items for the user is projected based on items liked by the user earlier. Then, the item with higher similarity would be recommended, irrespective of user preferences. In this scenario, the filtering method will not be able to recommend users unless new items are liked by a substantial number of users, known as the ***new-user problem*** [9]. Academic content-based recommender systems [10 - 14] recommend citations and research articles (Table **1**). Philip et al. [15] represented research

CHAPTER 6

Movie Recommendations

Anukampa Behera[1], **Chhabi Rani Panigrahi**[2,*], **Abhishek Mishra**[3], **Bibudhendu Pati**[2] and **Sumit Mitra**[4]

[1] *Department of Computer Science & Engineering, ITER, S'O'A (Deemed to be) University, Bhubaneswar, India*

[2] *Department of Computer Science, Rama Devi Women's University, Odisha, India*

[3] *Indian Institute of Technology, Bhubaneswar, India*

[4] *Managing Partner, Citizen, Odisha, India*

Abstract: Recently, most retail-based and e-commerce companies have been using recommender systems aggressively. It retains a customer's interest by giving exclusive offers on personalized preferences. The primary purpose of a recommender system is to get at an increase in sales by providing an enriched experience to the customer. With the emergence of many video streaming services like Netflix, Hotstar, and amazon prime video, the dependency on movie recommendation systems has increased. It facilitates the users in faster search and easier access for shows matching their tastes and helps them choose what they are looking for without getting lost in the flood of available options. The user most often gets surprised by seeing an offer that they possibly would never have searched. The system is based on information retrieved and processed user preferences, ratings, likings, disliking, *etc*., to use this understanding to recommend the products. In this chapter, we have discussed the various popular algorithm used for the movie recommendation, along with an insight into the extensive use of models based on machine learning especially deep learning. The performance of different movie recommendation systems with a comparative analysis is also given to encourage further research in this area.

Keywords: Collaborative filtering, Content-based filtering, Demographic filtering, Hybrid recommender, Personalization, Similarity test, Video ranker.

INTRODUCTION

Towards the end of the second decade of the 21st century, the world that we humans live in, has changed by the touch of technology - sometimes directly or indirectly. In some way or the other, technology has become an inevitable part of our life. Cutting-edge technologies have devised numerous business opportunities

* **Corresponding author Chhabi Rani Panigrahi:** Department of Computer Science, Rama Devi Women's University, Odisha, India; E-mail: panigrahichhabi@rdwu.ac.in

Abhishek Majumder, Joy Lal Sarkar & Arindam Majumder (Eds.)

for organizations by helping them become more responsive, agile, and flexible to act smarter. We have started living more in the info-sphere created, which is coherently online, offline, analog, or digital. We spend a significant portion of our everyday time in the virtual world, remaining connected to the internet through numerous gadgets and devices. For instance, these can be our smartphones, computers, smart TVs, intelligent cameras, echo devices, or baby monitors [1]. Especially smartphones have gained enormous popularity turning our lives immensely dependent on them. Much crucial information like business details, products used, feedback, and ratings given, can be accessed *via* a mobile phone. Recommendation systems are developed for mining all the contextual user information like reviews, ranks made, votes cast, the number of times a video has been watched [2], *etc*.

A recommendation system is based on filtering preferences from a large number of user data. The system makes recommendations based on user priority challenging to be aware of his choice, mood, *etc*. There is a large range of areas where recommendation systems are used, such as e-commerce, retail, media, banking, telecom utilities, tourism, *etc*. An example can be recommending an itinerary to a tourist based on his point of interest. This can be achieved by comparing his/her local travel sequences with the travel sequences of other tourists who share similar points of interest using geo-tagged photographs [3]. In recommendation systems, an auto-analysis is done on the customer data to suggest customized products/services. These data can be collected implicitly from purchase history and browsing details, or they might be from explicit feedback and ratings provided by the user. To understand this from a layman's perspective, let us assume a user of Netflix is browsing through the available list of web series and movies and reading the details. Each time he/she clicks a link, an event is fired to make an entry in a database like “User X clicked Item Y 1 time”. Similarly, every click is tapped. The user details can be retrieved from an HTTP session or system cookies. Though these all have their own challenges, the data collected can still be used to identify someone's preference in a particular mood and time.

According to Khanian *et al*., a Recommender system (RS) is a tool that assists users by presenting services or products that are most likely of interest [4]. In the current situation, the impact of the recommender system on the industry as well as academia is enormous. A large variety of recommendation system applications are made available to users nowadays. Almost every online platform implements it seeing its increasing popularity and acceptance. The content may vary from implementation to implementation. They are applied differently for products on an e-commerce website to stories on social media, and similarly, for films and podcasts, it is different from books and videos. It can filter data about user behavior, like Facebook suggests several stories based on the past clicks that we have made. Sometimes the group behavior helps in improvising the system. For instance, suppose it is observed in Amazon that whoever purchases cell phones is also looking for a hands-free speaker. Then hands-free speakers can be recommended to a customer once he adds a phone to the cart. Recently, users have started trusting the recommendations due to the technological advancements that have made the system perform better [5]. People are not sticking to the services that are not so suitable or without the recommendation, which has compelled the organizations to make substantial improvements to the existing recommendation structures [6]. Catering to users' needs whose choices, likings, preferences, and disliking vary radically. Often, a single person's wish depends on many factors like mood, workload, recent activity, and season, which is really challenging and complicated. For example, our movie choice during an outing with family may not be the same when we sit alone to relax after a hectic day at work. So to perform better, every time the customer uses the application, the system needs to discover more and more details regarding him/her.

The remaining chapter is organized as follows. A brief overview of the movie recommendation system is provided in section II, followed by various categories of recommendation systems in section III. The related works done in designing a recommender system using multiple machine learning and deep learning technologies are discussed in section IV. Netflix is discussed as a case study in section V as one of the most successful implementations of the movie recommendation system. Section VI lists different evaluation parameters popularly used to measure the efficiency of a recommender system, followed by the conclusion in section VII.

MOVIE RECOMMENDATION SYSTEM

Storytelling has always been considered at the heart of human nature. The path-changing technological innovations have brought a fundamental paradigm shift in society which has bestowed the world with more engaging and richer stories [7]. From the age when our ancestors enjoyed stories sitting around the fire to reading stories from printed press media to watching movies and TV series, it has now reached a cutting edge. In this cutting edge, one can watch web series and films through smart phone using the technological revolution brought by the internet. A movie recommendation is considered one of the salient applications of mobile services. It is a potent tool for providing users with valuable and appropriate personalized movies and web series suggestions. With abundant information that is flooded in the application, the user may get confused, lost, and eventually lose interest in using the application. The movie recommender system works as a

CHAPTER 7

Sentiment Analysis for Movie Reviews

Balajee Maram[1,*], **Suneetha Merugula**[2] and **Santhosh Kumar Balan**[3]

[1] *Department of Computer Science & Engineering, GMR Institute of Technology, Rajam, Andhra Pradesh, India*

[2] *Department of Information Technology, GMR Institute of Technology, Rajam, Andhra Pradesh, India*

[3] *Department of Computer Science & Engineering, Guru Nanak Institute of Technology, Hyderabad, Telangana, India*

Abstract: The viewpoint, together with a feeling examination, helps in distinguishing the perspective and characterizing the survey as good, nonpartisan, or negative. The angle-based conclusion investigation incorporates pre-processing of surveys, extremity figuring, and distinguishing the perspective and execution assessment. Primary assets of constant assessment, Twitter, messages, and other interpersonal organizations have captivated the significant interests of the exploration business and network. Wistful analysis (supposition mining) of brief, casual writings *via* web-based media sums up feelings as good, unbiased, or negative expressions of the sentiment holder. We will utilize a viewpoint-based estimation examination by using profound learning calculations. This methodology receives language preparation procedures, rules, and dictionaries to address a few opinion examination difficulties and produce summed results. By utilizing deep learning, we can yield more exactness than Artificial Intelligence (AI) calculations.

Keywords: Comparative opinion mining, Sentiment analysis, Accuracy, Deep learning, Reviews.

INTRODUCTION

Successful data mining analysis has formed a range of strategies, tools, and algorithms for dealing with large numbers of knowledge to unravel or resolve real-world problems. With increasing internet and technology, people get the freedom to express their views. They can often write reviews on the issues they often see and use the form of feedback and opinions. Many people currently use the web and shop on it and will look for good things in the longer term.

* **Corresponding author Balajee Maram:** Department of Computer Science & Engineering, GMR Institute of Technology, Rajam, Andhra Pradesh, India; E-mail: balajee.m@gmrit.edu.in

Abhishek Majumder, Joy Lal Sarkar & Arindam Majumder (Eds.)

Customer reviews are now more of interest for the service providers or product providers because they contain the customer's views. Service providers face complex problems in identifying their customers' behavior or interests. As people have different blogs, IMDB, Twitter, and forums, data analysts need more attention to get the feelings from the reviews posted in the comments by the public. Special attention is required to draw the individuals' views to expand the data analyst business.

SENTIMENT ANALYSIS

With progress, sentimental analysis has shaped a range of techniques, applications, and calculations for dealing with vast amounts of information to solve issues in the real world. The critical reasons for the records mining process are to monitor enormous amounts of information, mine for rules, and increase informative data. With the development of the website and its innovation, individuals have a chance to communicate their opinions and interests. They can audit the issues they see and regularly use within the type of assumptions or criticism. The present specialist organizations or item suppliers are keeners on the surveys of their clients since they contain the client's assessment and his/her enthusiasm about that item or administration. Specialist organizations are confronted with testing issues in discovering the conduct or enthusiasm of their client. Since individuals have various websites, IMDB, and Twitter, gathering conversations and information examination require more consideration of the audits that the individuals posted in the remarks.

As essential assets of ongoing assessment, Twitter, messages, and other informal organizations have drawn the significant interest of researchers and business organizations. Thoughtful analysis (feeling mining) of brief, casual writings *via* online media sums up conclusions as a good, impartial, or negative proclamation of the sentiment holder. A million quantities of tweets are made day by day on diverse issues. Phonetic adaptability in articulation and assorted topical variety in the content are two main difficulties in inspecting tweets. Various twitter estimation analyzers rely upon different supposition dictionaries to channel highlights to classifier models or to decide assessment scores. The prerequisite for a detailed examination is raised because of the expanded imperative of breaking down and organizing the covered data, which comes from online media.

By utilizing a phrase-level notion investigation, we can tackle the issues of information investigators while settling on significant choices. To get the notion, we have to make a machine to learn, which should be possible through administered and unaided learning. Classmarks of clear information is predicted by grouping. Class signs are predestined, and the classifier model is manufactured. We train the classifier for some information and then conduct new information tests. This classifier tries to determine the relationship between attributions and orders that rely on the most extreme probability relationship. Naive-Bayes, and Logistic Regression, are administered text feature calculations.

Honest Bayes-type computations are the extra adaptable estimations to order documents subject to repeated expressions. For training the classifier, simple Bayes classifiers consider a growing number of features. Bayes' principle of suspicion freedom is followed by gullible Bayes estimation. Another classifier in this area is Strategic Regression. In comparison to Naive Bayes, it is a tolerable classifier. Key Regression gives an excellent course of action that is twofold: yes or no. A machine has been built up considering the previously mentioned factors. This machine is used to aggregate at the end using feature-level requests asking positive and negative points of view. Naive Bayes and Logistic Regression are used to evaluate the testing data. Using record and sentence-level notion mining strategies, customers are depleted with any substance. While in perspective-based notion mining, by using the furthest point, customers get the most energizing sections. The goal is to discover energizing parts or remember chiefly subordinate for the furthest point and relative criticalness.

LITERATURE SURVEY

The different Natural Language Processing (NLP) techniques for evaluation examination were introduced by Surya *et al*. [1]. There are many well-known methods for pre-preparing printed material. Tokenization, state separation, grammatical form marking, and parsing are some pre-processing ventures needed for organizing printed content and separating capabilities. For various levels and situations, there are a few methodologies for feeling digging. There are five levels for which the conclusion mining system can be applied, and different methods can be applied for each level. A few devices are utilized for tokenization and division. The ground-breaking toolboxes for the NLP are NLTK, OpenNLP, CoreNLP, and so on. Feeling mining and opinion assessment intend to remove the notion of direction of writing. We can order conclusion mining into three areas: documental level, sentence level, and fine-grained level. Record stage assessment mining is a strategy of removing the general notion polarities of given data, comprehensive of online diaries, item reviews, and tweets. High-caliber grained sentiment mining is proposed to find data in stubborn writings, which gets super interests. Supposition mining on the sentence level is much the same as that on the record level since sentences can appear as snappy documents. A few open issues exist in utilizing NLP procedures in feeling mining [1].

CHAPTER 8

A Movie Recommender System with Collaborative and Content Filtering

Anupama Angadi[1], **Padmaja Poosapati**[1], **Satya Keerthi Gorripati**[2] and **Balajee Maram**[3,*]

[1] *Department of Information Technology, Anil Neerukonda Institute of Technology & Sciences, Visakhapatnam, Andhra Pradesh, India*

[2] *Department of Computer Science and Engineering, Gayatri Vidya Parishad College of Engineering (Autonomous), Visakhapatnam, Andhra Pradesh, India*

[3] *Department of Computer Science and Engineering, GMR Institute of Technology, Rajam, Andhra Pradesh, India*

Abstract: In the Internet age, we perceive the use of recommender systems all around us. The exponential growth of information from intelligent devices on the internet creates confusion for customers to pick a preferred product. Suggestions are a noble way to guide shoppers to discover fascinating products to impress customers. These recommender systems influence our browsing or watching or listening, searching patterns, and guess what customers might like in the future based on our patterns. For instance, a customer searching for baby products recommend diapers. Two significant categories of recommender systems exist, which are either collaborative or content filtering. The core of the recommender system resides in filtering similar users (or products). We address the introduction, existing works focusing on collaborative and content recommender filters, and their merits and demerits. Later, we classify types therein and thoroughly discuss similarity metrics used to filter neighborhood and evaluation measures used in the recommender system.

Keywords: Filtering, Movie recommendation, Similarity metrics, Quality metrics.

INTRODUCTION

Today, recommendation systems (RS) are present in our everyday lives. RS makes suggestions by personalizing customers' browsing experience, leading us to which films to watch (Amazon Prime), which series to watch (Youtube), what

* **Corresponding author Balajee Maram:** Department of Computer Science and Engineering, GMR Institute of Technology, Rajam, Andhra Pradesh, India; E-mail: balajee.@gmrit.edu.in

Abhishek Majumder, Joy Lal Sarkar & Arindam Majumder (Eds.)

to purchase (Big Basket), where to eat (RestroApp), whom to pair with (Tinder), *etc.* Many marketing strategies showed that personalization enhances customer satisfaction, which raises revenue [1-5]. As an outcome, web services try to capture comprehensive user profiles to understand customer needs, preferences, and buying patterns. Building intelligent and intuitive RS by observing the history of customers' recommendations gives deep insights into the fundamentals of RS [6 - 12].

This section serves as the beginning topic and establishes the fundamental notions and terminology for the rest of this chapter. Next, the background and overall outline of the recommender system are presented [13 - 18]. After that, categories of recommender systems, such as collaborative and content filters, are discussed. A section also discusses the significance of filtering similar users or products using various methods, such as weighted aggregate, Jaccard, and Pearson metrics. Finally, a quick outline of evaluation metrics is described and concludes the chapter.

The rest of the chapter is arranged as follows: In sections 2 and 3, we discuss the required related work and background. In sections 4 and 5, we present the types and algorithms of RS. Identifying similarities and evaluation metrics are provided in Section 6 and Section 7, respectively. We conclude the framework in Section 8 with possible future work.

RELATED WORK

Since the RS was established, product recommendations have become famous for investigators who have studied many techniques to overcome the drawbacks of the traditional approach. We brief the outlook of RS approaches from three perspectives: limitations, proposals of new similarity metrics, and accuracy.

Limitations

RS is the key to success in the digital world. However, it suffers from information overload, computational cost, cold-start, and scalability. Greg Linden [2] proposed a scalable RS that stores all radical changes and can operate in challenging domains. They described that traditional CF is expensive in computing, *i.e.*, O(MN) is the worst case; it has been reduced to O(M+N), where M and N denote the number of customers and items. Gilda Moradi Dakhel [4] claimed CF is successful but suffers from high computation and poor accuracy. They proposed a clustering-based approach based on Minkowski distance to resolve this issue. In the study [6], the authors reviewed these obstacles and suggested possible extensions for the new generation RS.

Proposals of a New Similarity Metrics

The traditional similarity metrics, such as Pearson, Cosine, Jaccard, and mean square deviation, suffer from few co-ratings and cold-start items [12, 18]. Most users show less interest in providing reviews leads to few co-ratings and low accurate similarity. For this reason, new similarity metrics were proposed to alleviate these drawbacks. Patra [12] presented a BCF metric that considers all ratings of an active user not to depend on co-ratings, unlike traditional methods. In the study [9], authors measured the proportion of co-ratings among users and used mean and variance to relate user preferences.

Accuracy

These CF algorithms' performance was evaluated using traditional precision and Mean Absolute Error metrics. McLaughlin [3] proposed a qualitative metric called the Belief Distributed algorithm and showed improvement to previous measures. Table **1** shows a brief review of studies on the recommender system.

Table 1. A Literature review on the recommender system.

Category	Techniques	Advantages	Limitations	Domain	Ref
Item-to-item collaborative filtering	Ranking each item	Scales well independent of the number of users (or items)	Complex and Expensive	Amazon.com	[2]
Collaborative filtering	Belief Distribution Algorithm	Preserves the quality of the nearest-neighbor algorithm	Increased Precision	MovieLens	[3]
Item-based recommender	KNN	Item-based are better than user-based algorithms	Can't perform well with sparse data	MovieLens	[4]
Collaborative Filtering	k-means clustering	More accurate than the traditional approach	More time consuming	MovieLens	[5]
Collaborative Filtering	Memory and model-based	Resolve drawbacks of the CF method	Difficult to implement	Books, News Articles	[6]
Content Filtering	Meta-data	Overcome the new-item problem	No better result	MovieLens, IMDB	[7]
Content Filtering	LDA	Scalable	Difficult to implement	MOOC's courses	[8]
Collaborative Filtering	Similarity metric	Accurate	Mathematical non-linear function	MovieLens	[9]
Collaborative Filtering	Jaccard similarity metric	Improved prediction quality measures	Difficulty in obtaining sufficient data	Netflix, MovieLens	[10]

CHAPTER 9

An Introduction to Various Parameters of the Point of Interest

Shreya Roy[1,*], **Abhishek Majumder**[1,*] and **Joy Lal Sarkar**[1,*]

[1] *Mobile Computing Lab, Department of Computer Science and Engineering, Tripura University, Tripura, India*

Abstract: Point-of-Interest (POI) recommendation helps to find new places for users to visit, along with the popularity of locations. Recommendation of POI is the most important in location-based social networks (LBSNs). This paper discusses different parameters that significantly impact the POI recommendation process and make the prediction much more accurate. A comprehensive review of a few research works and the methodologies employed for POI recommendation have been presented. POI recommendation techniques have been classified based on many, such as the interest of the tourist in particular POI, popularity of the POI, weather conditions, *etc.* A summary of related research work is presented for each category, along with their respective drawbacks. Finally, the possible directions toward future work in this area are included, along with the conclusion.

Keywords: Check-in, Point-of-Interest, Popularity, Recommendation, User interest, Weather details.

INTRODUCTION

Location-based Social Networks (LBSNs) have gathered huge followers recently [1 - 4]. LBSNs helps the users to connect and share their experience of visiting different Point of Interest (POI) which allows capturing their interests so that they can recommend those places that have not been seen yet [5 - 9]. Recommending locations to different users based on their liking is something challenging. The recommendation process requires many parameters to achieve an enhanced and well-structured recommendation list. A good observation of essential factors of Point of Interest improves the accuracy of recommending different locations to users. POI recommendation is a service aiming to recommend new POIs to multiple users who have not visited them before [10 - 14]. Several studies have

* **Corresponding authors Shreya Roy, Abhishek Majumder & Joy Lal Sarkar:** Mobile Computing Lab, Department of Computer Science and Engineering, Tripura University, Tripura, India; E-mails: royshreya654@gmail.com, abhi2012@gmail.com, joylalsarkar@gmail.com

Abhishek Majumder, Joy Lal Sarkar & Arindam Majumder (Eds.)

been done regarding better and more accurate recommendations [12, 24]. A Point Of Interest (POI) basically proposes any personalized recommendation offered to the user based on user-defined data or open data. There is a well-known psychological fact that humans always tend to behave consistently [1, 2], which effectively helps to learn and predict the patterns of various human behaviors. Recently Location-Based Social Networks (LBSNs) are very popular and possess rapid development, just like smart mobile devices and web connections, with the help of users sharing their locations. There are some typical LSBNs available, such as Foursquare [3], Yelp [3], Facebook, Geolife, and Gowalla [3], *etc.* that help users to built connections, upload photos, and share locations *via* check-in data [3]. The LBSNs need to have rich information and be very much prompt about user preferences to recommend new places to users that the user may be interested in visiting. The recommendation system has been widely studied and adopted by many e-commerce sites like Amazon, Netflix, and Facebook. Nowadays, POI is one trend.

In this paper, the main focus is given to the different influential factors of POI recommendation and how they are effective in recommending different locations to the users, along with some research related to these factors, which are discussed further.

IMPACT OF VARIOUS PARAMETERS ON POI RECOMMENDATION

Productive Point-of-Interest (POI) recommendation of un-visited locations for users is a real challenge for LBSNs. Recommendation of POI depends on a few influential factors, mainly dealing with user check-in activity. Due to many physical constraints and heterogeneous information, spatial and temporal properties are achieved. The demonstration of a few influential factors is given below.

Users' Interest-Based Recommendation

The user's interest in specific POI has been extensively explored in the recommendation systems for better prediction. Based on their check-in to different locations, the recommendation is made. Many researchers have done so much work regarding this factor.

Liao et al. [4] proposed a new strategy with tensor factorization to achieve accurate recommendations. First of all, Latent Dirichlet Allocation(LDA) [4] uses all the users' comments that extract topic information and generic topic probability distribution of POI. All the check-in information is divided into slices corresponding to every hour of the day. The LDA model used here is language-based. It mainly uses a natural language model that extracts topics and provides the topic distribution of every POI.

Fig. (**1**) shows that time-aware user-topic distribution is found by fetching POI-topic distribution and check-in and combining them. A user-topic-time mode is conducted by combining the topic distribution of visited POI. Finally, an algorithm named higher order singular value decomposition [4] was used to get more dense preferences by applying decomposition of the third order tensor.

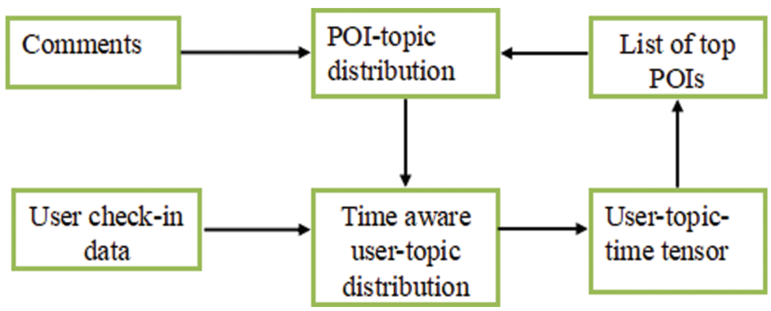

Fig. (1). POI recommendation framework following [4].

The aim of *Chen et al*. [5] is to obtain the list of POIs for the target users, considering geographical [5] as well as temporal influences [5]. A probabilistic method was defined to achieve the purpose initially. It detects the user's spatial orientation [5] by the weights of POIs that those users visit. After that, users' temporal preferences are detected to obtain collaborative filtering. Finally, by combining these two methods, the POI recommendation was performed.

An absolute user spatial orientation observation [5] is achieved in the following way to show that users can be oriented toward familiar or unfamiliar check-ins.

$$u_i^s = \frac{\sum_{l_j \in L_i} \left| l_{i,j}^o \right|}{|L_i|}$$

Where, $l_{i,j}^o$ is location's absolute POI spatial orientation [5] limited to the user's visits. Then, for each visited POIs by the users, the absolute rate of spatial familiar or unfamiliar check-ins deviation is calculated as,

$$l_i^s = \frac{\sum_{u_i \in U_j} \left| l_{i,j}^f - l_{i,j}^u \right|}{\left| U_j \right|}$$

CHAPTER 10

Mobile Tourism Recommendation System for Visually Disabled

Pooja Selvarajan[1], **Poovizhi Selvan**[1,*], **Vidhushavarshini Sureshkumar**[1] and **Sathiyabhama Balasubramaniam**[1]

[1] *Department of Computer Science and Engineering, Sona College of Technology, Tamilnadu, India*

Abstract: Mobile Tourism Recommendation System recommends to a tourist the best attractions in a particular place according to his preferences, profile and interest. First, a Recommender system offers a list of the city places likely to interest the user. This list estimates the user demographic classification, likes in former trips, and preferences for the current visit. Second, a planning module schedules the list of recommended places according to their characteristics and user limitations. The planning system decides how and when to perform the recommended activities. For implementing these recommender methods, we have applied different machine learning algorithms, which are the K-nearest neighbors (K-NN) for both Clean Boot (CB) and Consolidation Function (CF) and the decision tree for all Data Framing (DF). Thus, executing a recommendation system for tourists helps them with user-friendly planning. Blind people can also use this. This application provides complete voice assistance for easy navigation *via* a simple button click. Vibratory and voice feedback is provided for accurate crash alerts for visually challenged people. The application extracts its smartness by incorporating Android and Internet of Things (IoT) support. Since blind-supported applications and devices are more expensive and many blinds can not afford them, we aim to put forth a novel, low cost and reliable approach to help the blind explore the possibilities and power of smartphone technology in navigation. We additionally expect to find the static variables that should be tended to, food, tidiness, and opening times, and valuable to suggest a tourist place depending on the travel history of the client. In this investigation, we propose a cross-planning table methodology depending on the area's prevalence, appraisals, idle points, and conclusion. A targeted work for proposal streamlining is defined as dependent on these mappings. Our outcomes show that the consolidated highlights of Latent Dirichlet Allocation (LDA), Support vector machines (SVM), appraisals, and cross mappings are helpful for upgraded execution. The fundamental motivation of this study was to help businesses related to tourism.

* **Corresponding author Poovizhi Selvan:** Department of Computer Science and Engineering, Sona College of Technology, Tamilnadu, India; E-mail: poovizhi.19cse@sonatech.ac.in

Abhishek Majumder, Joy Lal Sarkar & Arindam Majumder (Eds.)

Keywords: Requirement Portal, Hybrid Recommendation System, Personalized Recommendation System, Deep map, KNN Algorithm, Ultrasonic Sensor.

INTRODUCTION

The tourism industry has always welcomed new technologies. With the usage of the web and electronics, mobile tourism (e-tourism) has grown significantly. At the same time, tourists play a more active role in the process of content production. They thoroughly publish information using web technologies, such as social networks, blogs, and wikis. The users also post dynamic content about visited destinations or relevant information about their visit, which can be helpful to other tourists [1 - 3]. Extending the notion of electronic tourism to meet the vision of tourist services and provisioning of migrant users with no spatial-temporal limitations is expected to become a reality within the next few years. In this context, the 'Mobile Tourism' field has emerged, wherein tourist information and services are accessed *via* mobile devices. Blind mobility is one of the main challenges that are faced worldwide. The number of visually challenged people is around 285 million, and 39 million are blind. Since blind support devices are more expensive, many blind people can not afford them. This work also aims to develop a cheap blind guidance system for developing countries. This chapter aims to develop a low-cost intelligent system for guiding the visually challenged by providing information about the environmental scenario of static and dynamic objects around them. The main functions of this system are path indication and environment recognition. The system is provided with a smartphone and a small embedded circuitry. Smartphone software is developed to recognize the destination place through voice command and draw a route from the current position to the desired destination [4]. Ultrasonic sensors will provide information about obstacles if they are within the limit range. Ultrasonic sensors are connected to a microcontroller circuit, and a microcontroller timer module is used to regulate the output of the sensor and send them to a smartphone *via* Bluetooth. Mobile tourism represents a recent trend in the tourism field that involves the use of tourist applications offering services and tours with multimedia content executed on electronic mobile [5].

The tourism Recommendation System is to enhance tourist decision-making. The tourist needs to understand how the system generates recommendations [6, 7]. It reduces users' efforts and preserves their privacy. The main purpose of this application is to develop a recommendation system based on the user's enhanced profiles [8]. These profiles will be composed of functionality levels regarding accessibility issues. Locomotion, vision, and phobias like acrophobia, agoraphobia, and claustrophobia are some accessibility issues. The user's basic information, like age, gender, and nationality, is also provided by the user profiles. This research assumes a vital role in tourism recommendation systems and other areas where individual user knowledge is a key factor. One of the most

important goals of this proposal is to fulfill the user's needs in coexistence with respect for their physical and psychological limitations [8, 9].

Our primary commitments are summed up as follows:

(a) An on-location travel conduct information gathering technique is intended to detect nearby travel conduct under indoor consequently and outside the travel industry situations. It considers travelers' cell phones and Bluetooth Low Energy (BLE) guides.

(b) Tourist-Behaviour Prefix Span calculation is proposed to create successive travel courses effectively depending on recorded Tourist-Behaviour design successions.

(c) Travel course positioning strategy is proposed to suggest specific travel courses as indicated by the questioning vacationer's profile and requirements. It guarantees the course esteem and reasonableness of the final travel courses.

The centralization of the analysis has been complemented by the growing popularity of portability and worldwide innovation. A suitable traveler should help sightseers with their planned schedule. An approach has been proposed in this work to suggest a schedule plan for travelers. When sightseers travel to an unfamiliar city or a country, they may have a fascination with specific spots they need to visit. Initially, different travel spots have been clustered depending on the geographic area and inclinations of the travelers utilizing the K-means clustering algorithm. The following stage is to track down the ideal itinerary for each group. For this reason, covetous and 2-select calculations are used in this work. In this way, a perfect schedule plan is suggested for travelers.

PROPOSED WORK

Recommender systems map user needs and constraints through algorithms and convert them into product selections. The framework of the proposed technique has been presented in Fig. (**1**).

Recommendation Systems

In this section, some of the existing recommendation systems have been presented along with their capabilities:

CHAPTER 11

Point of Interest Recommendation *via* Tensor Factorization

Shreya Roy[1,*], **Abhishek Majumder**[1] and **Joy Lal Sarkar**[1]

[1] *Mobile Computing Lab, Department of Computer Science and Engineering, Tripura University, Tripura, India*

Abstract: In the recent era, recommendation systems have marked their footsteps and have changed the way of the travel industry. The recommendation system deals with massive amounts of data to identify users' interests, making the location search easier. Many methods have been used so far for making predictions much more desirable regarding users' interests by collecting Information from a large set of other users. The main objective of this paper is to show various methods and techniques used for generating recommendations. These recommendation processes are classified into different forms, such as traditional methods and tensor-based methods. A brief review of these methods was described with the help of some challenges faced by the recommendation system. Apart from that, the advantages and disadvantages are discussed, along with the highlights of future directions.

Keywords: Point of Interest, Tensor factorization, Recommendation, Collaborative Filtering, Check-in, Tucker decomposition, Preference.

INTRODUCTION

Locations are critical in case tour and travel planning, selecting proper locations, and scheduling them accordingly [1 - 4]. Finding the best location and recommending it to the user is known as a location recommendation. Countries' recommendation systems are established based on different choices and interests of users over cities.

Point Of Interest (POI) recommendation system offers personalized recommendations to the user based on user-defined data or open data. It is pretty much well-known that human behavior is consistent [5, 6], which makes learning and predicting the patterns of human behaviors much easier [7 - 19]. With the rapid blooming of intelligent mobile devices and web connections, Location-Based

* **Corresponding author Shreya Roy**: Mobile Computing Lab, Department of Computer Science and Engineering, Tripura University, Tripura, India; E-mail: royshreya654@gmail.com

Abhishek Majumder, Joy Lal Sarkar & Arindam Majumder (Eds.)

Social Networks (LBSNs) help users share their locations [20 - 23]. There are some typical LSBNs available. Among them, Foursquare [24], Yelp [25], Facebook, Geolife, Gowalla [25], *etc*., help the user to build connections, upload photos, and share locations *via* check-in data [23].

The LBSNs need to have rich Information and be very much prompt about user preferences to recommend new places to the users they may be interested in visiting [26 - 34]. The recommendation system has been widely adopted by many e-commerce sites, such as Amazon, Netflix, Facebook, *etc*. Nowadays, POI is one trend. The POI recommendation needs accurate and prompt results. There are some factors where the POI recommendation system differs from traditional recommendation systems as it provides some unique characteristics.

According to Tobler e*t al.,* the First Law of Geography states, "*Everything is related to everything else, but near things are more related than distant things*" [19]. It simply implies that the user prefers nearby places rather than distant places [35 - 44]. Hence, it can be seen that POI recommendation is much more effective in predicting user preferences than traditional recommendation techniques, as it uses geographical influence to recommend new places to the user.

In early times traditional recommendation methods take the user ratings on an item or place, which are further converted into an item rating matrix [29]. The ratings are mainly based on numerical range, *e.g*., 1 to 5. A score of 5 is the higher rating and is considered the best satisfactory result. Unlike such recommendations, user preferences are considered using the check-in, which forms the check-in frequency matrix of user location [29]. The frequency range is larger than ratings, but the sparsity [29] (*e.g*., 90% sparse means that 90% of its cells are either not filled with data or are zeros) of the matrix is very high. It makes POI recommendations very challenging.

Apart from that, social influence is also another way to know user preferences. Most users tend to follow their family and peers' choice. Hence, the traditional recommendation system takes the user preferences and ratings to improvise the recommendation which is found in the previous studies [45, 8]. In the case of POI recommendations based on early research studies, users share a significantly less common interest, which means social influence is less effective for check-in behavior.

Influential Factors of POI Eecommendation

Different user modeling approaches or algorithms are available for POI recommendation according to different types of LBSN data.

Pure Check-in Based POI Recommendations

Traditional recommendation systems are used to gather ratings given by users on different items or places that are not available in the LBSNs [5]. In the case of POI recommendation, the frequency of check-in to different locations is considered as the user preference. With this available check-in data. The user visits only a few locations in LBSNs, so the user data is encoded into a sparse matrix [14].

With all those Information, the recommendation approaches are employed for POI recommendation. The two most popular POI recommendation processes are User-based collaborative filtering [28] and item-based collaborative filtering [28]. Collaborative Filtering (CF) [1, 29] based recommendations are achieved by Matrix Factorization(MF) [21]. In the case of user-based POI recommendations, the similar taste of similar users is considered, making POI recommendations effective.

While in the case of item-based POI recommendations, users with similar kinds of POIs are considered. The user's check-in history can be considered as binary values such as '1' or '0'. The value of the matrix will be 1. If the user visits a particular location based on different categories, *e.g*., restaurants, cinema halls, residences, shops, *etc*., the value will be '1'; otherwise, '0'. If the check-in result is shown like this, it may ignore the frequency of check-in. Few users visit their preferred location. Therefore, many entries in the matrix are left with '0's. It creates data sparsity, and to overcome such a problem, both the user-based and item-based POI approaches are applied. Various studies [29] found that user-based POI recommendation is better than item-based one, which leads to inaccurate item similarity as compared with user similarity. There is model-based collaborative filtering [29] which can be adopted for POI recommendation apart from a few memory-based collaborative filtering [28]. After being utilized by Google and Netflix, the MF, especially the CF modeling approach, have gained popularity for POI recommendation as it is much more effective in dealing with large user-item rating matrix [14]. Moreover, the regularization MF-based POI recommendation proposed by B. Berjani and T. Strufe [2] dealt with the lack of explicit rating. Another POI recommendation presented by Wang *et al.* [22] decomposes two low-rank latent feature matrices just for modeling the importance of 'venue semantics' in the check-in behavior of the user. The knowledge about content and context (*e.g*., POI category, user context, sentiment indication, and time stamp) in LBSNs also helps to publish other types of characteristics of the check-in behavior of several users. The advantages of using such approaches are to reduce dimension and minimize data sparsity. The drawback of this approach is

CHAPTER 12

Exploring the Usage of Data Science Techniques for Assessment and Prediction of Fashion Retail - A Case Study Approach

Dillip Rout[1,*]

[1] *Department of Computer Science and Engineering, Centurion University of Technology and Management, Odisha, India*

Abstract: In this article, the insights of a garment retail store have been studied with respect to the attributes of the dresses and sales information. Mention that each dress in fashion retail has several attributes or features. These features play a critical role in the selection of consumers or customers. This study tries to establish the relationship among these features by which the importance of the attributes is evaluated concerning sales. Furthermore, this paper tries to automate the process of the recommendation of the dresses by using these attributes. It is merely a binary classification but useful for retail sales. Moreover, the demand for sales is estimated over a period. All these objectives are achieved through using one or more data science techniques. The case study shows that the algorithms of data science are helpful in the decision-making of fashion retail.

Keywords: Categorical Features, Data Wrangling, Feature Engineering, Market Basket Analysis, Time Series Forecasting.

INTRODUCTION

The fashion and garment industry is expanding very fast in this modern era. The underlying reason for the rapid growth is the diversity in demand which in turn has an expeditious change because of the trend in the choice of customers [1]. Consequently, the business model evolves over time, needing regular revisions in order to match the swinging demand and choice [2, 3]. In such a scenario, the basic thumb rule and/or general perception will not be feasible enough to ensure a smooth operation of trading by tackling this issue. Hence, some of the manufacturing units and retail outlets of the fashion industry have already automated their businesses. Nevertheless, automation is a typical and critical aim

[*] **Corresponding author Dillip Rout:** Department of Computer Science and Engi- neering, Centurion University of Technology and Management, Odisha, India, Email- dil- lip.rout@cutm.ac.in

Abhishek Majumder, Joy Lal Sarkar & Arindam Majumder (Eds.)

to achieve as it requires an ample understanding of the underneath business logic, cost, attributes of fashion, demand, choice of customers, *etc* [4]. Therefore, it is not surprising that many of these industries are yet to be automated, and the already automated industries struggle to maintain the consistency of profit. The major bottlenecks to such a circumstance are the characteristics of the industry, which differ from case to case apart from the varying demand and choice of customers [5]. Thus, understanding the data is vital in this situation, and the formulation of models is a necessity for the fashion retail industry, possibly at an individual level.

In this regard, the usage of data science techniques is explored in this paper for the automation of a fashion retail outlet. A dataset of retail information is given, which contains attributes or features and sales of a set of garments. The primary aim is to find the importance of garments that influences sales, listing out the major features which led to higher sales. Furthermore, the user ratings are inspected to figure out the importance of features on total sales. In addition, the analysis is also focused on forecasting the future demand for each garment of retail based on history. Each of these questions is dealt with one or multiple machine learning algorithms, such as Linear Regression, Logistic Regression, Decision Tree, Random Forest, *etc*. Mention that one algorithm may not be sufficient to find a definite solution as it is a data-driven approach, hence, a combination of algorithms is used to recommend the best solution. Besides measuring the features, the algorithms are also evaluated as fitting with respect to the objectives and data. All the applied algorithms are demonstrated through a case study which consists of 15 features, including attributes and sales data for 500 garments.

The remainder of the paper is organized as follows. The rest of this section has a literature review followed by the list of objectives of the study. Then in Section 2, a theoretical background of the applied machine learning techniques is discussed. Next, the article is described through the stated case study in Section 3. Lastly, the concluding remarks are presented in Section 4.

PREVIOUS WORKS

In this subsection, a review of previous works is presented. This paper focuses on fashion or garment retail, however, there are very few papers which has the same theme. Hence, the review is extended to the articles which address general retail problems. The literature shows that the retail industry is studied in three facets, namely, the behavior of consumers, decision-making, and demand forecasting. Most of the literature is focused on demand estimation for future production, whereas a few of the papers state the other facets. Here, a comprehensive review is presented for all three facets as follows.

The behavior of the consumers plays a crucial role in determining the need for production for the retail industry [1 - 5]. Note that the behavior spreads over both offline and online consumer behavior. In general, several classifiers are used to fit consumer behavior and build up a recommended system. For instance, the decision table classifier provides the highest accuracy level for consumer behavior for online shopping data [6]. On the other hand, the filtered classifier depicts the lowest accuracy in showing the behavior of the consumers. Finally, a recommended system is suggested by the authors by which consumers can easily select their required items. Furthermore, the authors state that clustering techniques, association rules, random trees, forests, *etc.*, are also attempted in the past to know the behavior of the consumers. Another article reveals that a combination of sentimental analysis and neural networks gives better precision in fixing the price of the products [7]. Thus, it is particularly important to analyze the data of the products as well as the needs of the consumers in order to settle the price of the product. Furthermore, the purchase decision and situations where the consumers buy are studied with digital signage to predict the behavior [8]. In this case, the support vector machine is proven to result in the best output with high accuracy. Differently, the usage of the Internet of Things in the application of smart stores is studied through the application of indoor positioning, augmented reality, facial recognition, and interactive display [9]. Furthermore, attitudes and subjective norms are found to be the key predictors for online fashion renting, which is found through confirmatory factor analysis and structural path analysis [5]. Ease of shopping is the most influencing factor in consumer behavior [10]. It is concluded that these applications improve the experience of consumers and their behavior of buying from offline stores. Overall, multiplexing technology has played a critical role in both online and offline shopping, and sales have increased.

Decision-making in retail management is extremely difficult due to the variety of features (characteristics) to be processed [11]. So, the dimensionality is reduced using correlation analysis first, and then a random forest classifier is applied to the lowly correlated features. Note that the features having high correlation are discarded as these are redundant while applying classification. Likewise, machine learning techniques are used with the help of non-parametric statistical tools to determine the performance of sales strategy [12]. It is found that discount offers are effective in increasing sales, possibly with the combined products. Nevertheless, there is a saturation point to the discount where the sales do not increase. The manufacturing of Zara is being shared by a couple of sub-continents because of its popularity [13]. Zara started its manufacturing in Europe but spread

CHAPTER 13

Data Analytics in Human Resource Recruitment and Selection

Sumi Kizhakke Valiyatra[1,*]

[1] *Institute of Management in Kerala, University of Kerala, Kerala 695034, India*

Abstract: Human resource data analytics are more important now than ever before. An increasing number of businesses are delving ever deeper into the data they collect about their employees, their success, and their well-being. Recruitment analytics can assist in making smarter, data-driven selection, recruiting, and sourcing decisions. This technology will scan hundreds of resumes at once to provide the best possible fit for a particular job opening. An organisation can submit automated emails with an interview appointment that automatically sinks with the work calendar using modern recruiting software. The organisation may use automated disqualification of unqualified applicants to automatically screen application forms and exclude candidates who aren't qualified. Effective recruitment is a mix of science and art. It necessitates the implementation of repeatable processes that produce consistent results.

Keywords: Recruiting analytics, Time to fill, Artificial Intelligence screening, Optimum Productivity Level, Selection Ratio, Attrition rate, Employee Life Cycle.

INTRODUCTION

Data analytics has gained much traction in this technological transformation age [1]. Data is obtained in its raw form, analysed according to a company's requirements, and then used for decision-making purposes. This procedure aids companies in growing and expanding their operations [2 - 5]. Data analysis is a constantly changing discipline that places a lot of emphasis on new predictive modelling techniques.

The methods and techniques used to analyse data in order to improve efficiency and profitability are known as data analytics [6 - 9]. To analyse different behavioural patterns, data is collected from various sources, cleaned, and classified. Depending on the organisation or entity, different methods and techni-

* **Corresponding author Sumi Kizhakke Valiyatra:** Institute of Management in Kerala, University of Kerala, Kerala 695034, India; E-mail: kvsumi001@gmail.com

Abhishek Majumder, Joy Lal Sarkar & Arindam Majumder (Eds.)

ques are used. Due to Data analytics, as a result of technological advancements, has a far-reaching influence on various fields of human development. Employee recruitment is a critical activity of every organisation. Any flaw in the selection process results in exorbitant costs [10, 11].

‘Data analytics in human resources is more important than ever before to an enterprise [2].' Although human resources managers continue to place a premium on people skills, a growing number of companies and non-profits are crunching numbers to evaluate everything from talent development and retention to efficiency and job structure [12 - 18]. The shift to analytics is both cost-effective and time-saving.

The amount of data an organisation will obtain to increase efficiency and output appears to be limitless. An increasing number of businesses are delving ever deeper into the data they collect about their employees, their success, and their well-being. Both companies and organisations prioritise retention and recruiting, which is also highly targeted for analytics. Analysis of large amounts of data helps us understand the huge volume of information our world generates [20].

RECRUITMENT ANALYTICS

In today's dynamic work market, recruitment analytics has a significant effect. Recruitment analytics can assist in making smarter, data-driven selection, recruiting, and sourcing decisions. It may be due to a mismatch between the job description and the actual position or a bad onboarding process, that a newly hired employee leaves within the first three months that are a case in point. Recruiting analytics can answer the following questions.

How much does it cost to employ someone for a job?

Which sourcing method yields the most qualified candidates?

What is the recruiting success rate?

It is important to be able to address these questions in order to enhance recruitment decision-making. Applicant tracking systems, satisfaction surveys and brand data. Customer Relationship Management systems, Human Resources Information System data, and data from work advertising platforms are all popular data sources for recruitment analytics.,

Procedure for Recruitment Analytics

Before you source, employ, and hire highly qualified talent, create a candidate persona to outline the talents, strengths, experience, and tendencies of the ideal candidate. A well-defined and accurate applicant persona will aid the talent team in tailoring its plans and approach to the talent that an organisation is trying to recruit. When it comes to high-level recruitment, this is particularly significant.

Using existing data from Applicant Tracking System, customer relationship management system and databases on prospective applicants, a Candidate persona must answer several main questions. Fig. (**1**) shows an example of a perfect persona Candidate Persona.

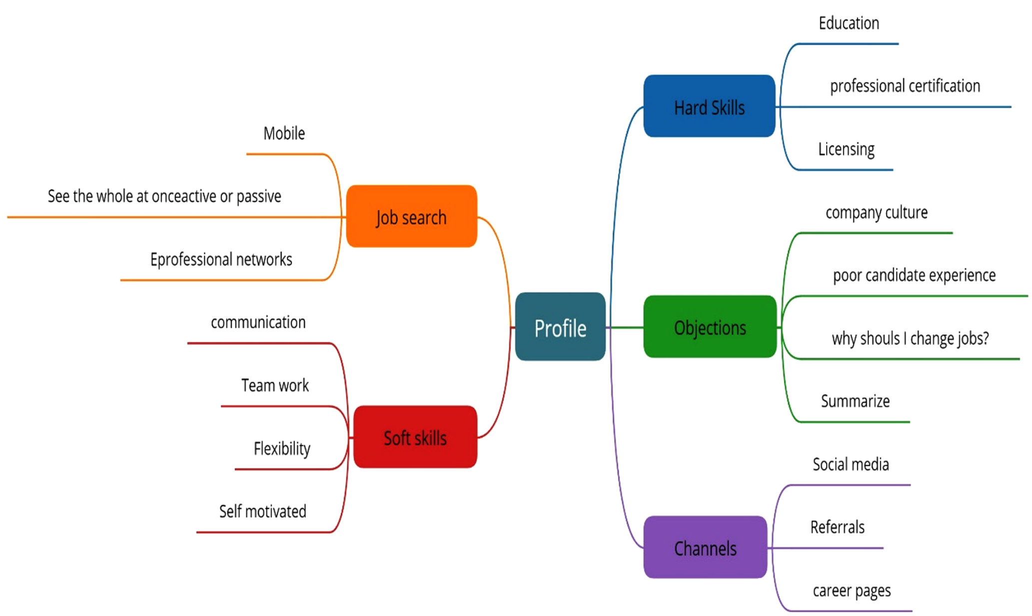

Fig. (1). Flow of a perfect persona.

OPERATIONAL REPORTING

Recruitment data is descriptive of nature. They are the well-known fundamental hiring metrics. The cost of recruiting, the source of recruitment, the number of prospective candidates per job opening, the quality ratio, the time required to fill, the time required to hire, and hiring manager satisfaction are a few examples of metrics.

Recruiting Metrics

Recruiting metrics are estimates used to track recruiting performance and success rates and improve the method of hiring candidates for a company. These criteria, when used correctly, aid in evaluating the recruitment process and determining if the organisation is hiring the right people. Recruiting metrics are the measures

CHAPTER 14

A Personalized Artificial Neural Network for Rice Crop Yield Prediction

Pundru Chandra Shaker Reddy[1,*], **Alladi Sureshbabu**[2], **Yadala Sucharitha**[2] and **Goddumarri Surya Narayana**[3]

[1] *Department of CSE, CMR College of Engineering & Technology, Hyderabad, TS, India*

[2] *Department of CSE, JNTUA College of Engineering, Ananthapur, Andhra Pradesh, India*

[3] *Department of CSE, Vardhaman College of Engineering, Hyderabad, TS, India*

Abstract: Early and accurate crop yield estimates at a local and national level are essential to oversee industry and trade planning and to mitigate the price hypotheses. The major challenge for farmers in the agricultural field is selecting an appropriate crop for planting. Crop selection is dependent on several factors like climate, soil nature, market, *etc.* Majorly, crop yield production depends on weather conditions and soil types. Yield anticipation is essential for farmers nowadays, which significantly adds to the appropriate yield selection for sowing. There needs to be a framework to recommend what type of crops to produce for farmers. It is essential and challenging to make the right farming decisions at a future steady cost and yield balance. This article proposes an Artificial Neural Network (ANN) model for rice crop yield prediction by utilizing weather parameters like rainfall, temperature, sunshine hours, and evapotranspiration. Generally, Default-ANN has only one hidden layer. But in this work, a Personalized Artificial Neural Network (PANN) approach has been designed by varying the number of hidden layers, the number of neurons, and the learning rate. P-ANN model accuracy is computed using R-Square (R2) and Percentage Forecast Error (PFE). Outcomes demonstrate that the P-ANN model performs precisely with a greater R2 and smaller PFE values than existing methods. For this research, the seasonal (Kharif & Rabi) weather dataset and rice yield data of Guntur district, Andhra Pradesh, India, from 1997-2014 have been used. Better paddy yield was forecasted by utilizing the P-ANN approach.

Keywords: Rice yield, Agriculture, Prediction, Crop, P-ANN.

INTRODUCTION

Rice is accountable for 80% of national food supplies, making crop yield prediction necessary to direct the worldwide commodity market. Basic to global

[*] **Corresponding author Pundru Chandra Shaker Reddy:** Department of CSE, CMR College of Engineering & Technology, Hyderabad, TS, India; E-mail: chandu.pundru@gmail.com

Abhishek Majumder, Joy Lal Sarkar & Arindam Majumder (Eds.)

food security, the agriculture industry is not short of embracing innovations to make work processes more effective and enhance overall harvest yield [1]. Precision Agriculture (PA) is used to portray the execution of innovative procedures in horticulture. Suppose agriculturists can forecast crop yield and underline regions in the field where the yield is less than anticipated. In that case, they can hypothetically endeavor to counter the falling procedures in yield at given areas in the fields. The yield forecast is one of the fundamental demanding problems faced in the agricultural field. Farmers' shortage of awareness about yield surplus, vulnerabilities in the climate conditions, seasonal precipitation, consumption of nutrition level of soils, fertilizer accessibility, price, pest restriction, post-yield damage, and other elements leads to a decline in the production of the crops [2].

Rice is the fundamental food crop, and being a tropical plant, it grows easily in a hot and damp atmosphere. Rice is primarily grown in rain-fed regions that get severe yearly precipitation. Farming is one of the fundamental parts of the national economy [3]. Agriculture, especially rice planting, has faced various issues for a reasonable length of time over decades; for example, phenomenal atmospheric conditions could bring about yield disappointment. Indeed, even after multiple improvements in agriculture, water management, pesticide, fertilization, and hybrid crop improvement, the advantage and long-term benefit of agriculture-associated sectors are yet treated unpredictably [4].

For a regular crop like rice, climate plays a significant role in choosing its yield. The basic contemplations for rice are undeniably an adequate and well-distributed rainfall and the accomplishment of certain perfect soils and temperatures. Also, the procedure of rice production includes various sectors, such as farming, transportation, milling, and marketing. In such a composite situation, early information on seasonal supplies may hold up selling policies and industry aggressiveness other than giving vital information to plan milling activities and rice shipments [5]. At last, precise predictions are made accessible to the public. Early caution is very helpful to avoid any instability in food prices and sudden production fall due to adverse conditions that frequently influence the principal food items. These contemplations prompted the development of a timely and accurate crop yield prediction framework. Yield forecasting is a complex and challenging farming task [6]. Each agriculturist is enthusiastic about knowing how much yield he is about to anticipate. Before this, yield prediction was made by considering the farmers' past knowledge of a particular crop. Different data analytic (Machine Learning) strategies are utilized to explore massive volumes of datasets and develop supportive classifications and patterns. The effect of observed regular climate conditions, for example, precipitation and temperature variability, on anticipated crop yield has endeavored an accurate crop management structure [7].

Fig. (**1**) demonstrates the contribution of farming to the national income and its share in exports for 60 years (1950-2010), and it explains that the percentage of agriculture in the National Income is regularly declining. Farming subsidies were just around 33% of the National Income against 54% in 1950-51. Similarly, the share of agricultural goods in exports has reduced from 52.5% in 1950-51 to just 14% in 2010-11 [8]. The production of crops should be increased to improve the agricultural contribution to the country's income. A conceivable reason behind the destitute contribution of the farming field to the GDP of Andhra Pradesh (AP) might be the absence of good crop management by agriculturists, just as by the government.

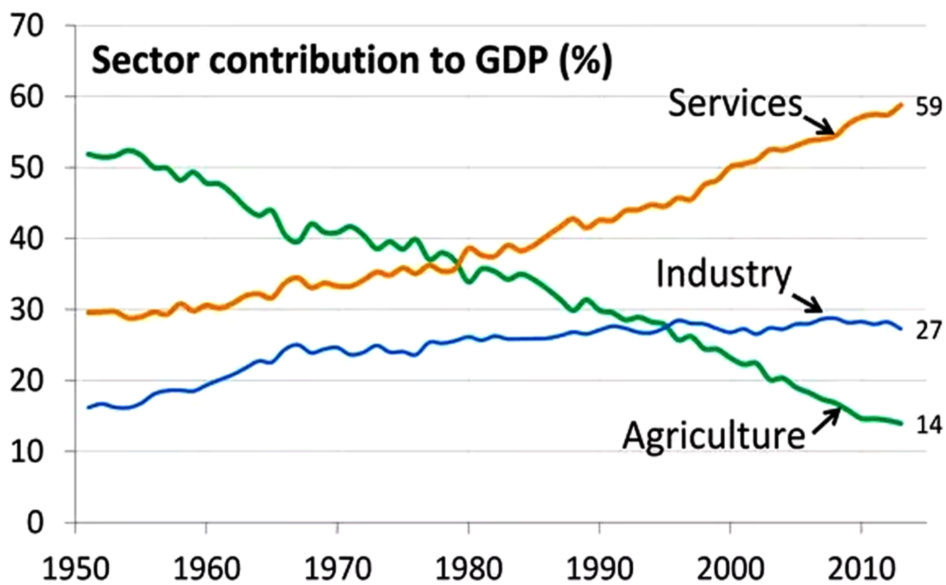

Fig. (1). Agricultural sector contribution to GDP (Courtesy: Wikimedia Commons).

Since the conventional approach of cultivation is proficient, there exists a glut or shortage of crops if a specific requirement is not achieved. The farmers don't know about the trade in the current farming economy. It results in the defeat of the agriculturists in the cultivation sector [9]. The conveyed causes in order of significance behind farmer suicides were climate conditions, low production costs, poor water management, and escalation in the price of farming. Various application areas have been developed with enhanced constraints and models in

SUBJECT INDEX

A

B

C

Abhishek Majumder, Joy Lal Sarkar & Arindam Majumder (Eds.)

D

E

F

G

H

I

J

K

L

M

N

O

P

R

S

T

U

V

W

Y

Abhishek Majumder

Dr Abhishek Majumder received his BE, MTech and PhD degree from National Institute of Technology Agartala, Tezpur University and Assam University respectively. He is currently working as Assistant Professor in the Department of Computer Science & Engineering, Tripura University. His research interests are Artificial intelligence, Recommendation System, IoT, and Mobile Computing. Some of his works published in Journals like Expert Systems and Applications, Electronic Commerce Research and Applications, IEEE/ACM Transactions on Computational Biology and Bioinformatics, Multimedia Tools and Applications, ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM) etc.

Joy Lal Sarkar

Dr. J. L. Sarkar is an Associate Professor, Department of Computer Science and Engineering, Techno College of Engineering, Agartala, Tripura, India. He received M.Tech and PhD degree from BPTU, Odisha, and Tripura University respectively. His research interests include Software-Defined Networks, Internet of Things, Recommendation Systems, Artificial Intelligence, Health Informatics. He is currently serving as a review editor in Frontiers in the Internet of Things. Dr. Sarkar is currently serving as an editor of International Journal of Advanced Intelligence Paradigms, Inderscience, International Journal of Computational Systems Engineering, Inderscience and section editor of Smart Tourism, Asia Pacific Academy of Science.

Arindam Majumder

Arindam Majumer is an Assistant Professor in the Mechanical Engineering Department at NIT, Agartala, India. He received his BE and ME degrees in Mechanical Engineering from NMU, Jalgaon and IIEST, Shibpur respectively. He obtained PhD degree from Jadavpur University, India in Mechanical Engineering. His current research interests are in the areas of advanced manufacturing systems, scheduling, soft computing techniques, robotics, and non-traditional machining. Some of his research articles have appeared in Knowledge-Based Systems, Computers & Industrial Engineering, Swarm and Evolutionary Computation, Neural Computing and Applications and Energy. Recently a patent of him is granted by Indian Patent Office.