A Survey on Data Mining Research Trends

Dhanalakshmi.D[1] IEEE Member, Dr.J.KomalaLakshmi[2] IEEE Member

[1]Research Scholar, Department of Computer Science, S.N.R SONS College (Autonomous),
Coimbatore, Tamil Nadu, India
dhanadurairaj@gmail.com

[2]Assistant Professor, Department of Computer Science, S.N.R SONS College (Autonomous),
Coimbatore, Tamil Nadu, India
jkomalalakshmi@gmail.com

Abstract: *Presently, a very large amount of data stored in databases is increasing at a tremendous speed. This growing need gives a view for a new research field called Knowledge Discovery in Databases (KDD) or Data Mining, which attract attention from researchers in from various fields which includes Database Design, Statistics, Pattern Recognition, Machine Learning, and Data Visualization et,. In this survey approximately 40 research papers were collected concerning various fields in data mining and discussed each and categorized them under the few areas and the trends was interpreted based on the area of research and applications.*

Keywords: Data mining, KDD and Machine Learning.

1. Introduction

Data mining an interdisciplinary subfield of computer science is the computational process of discovering patterns in large data sets involving methods at the intersection of artificial intelligence, machine learning, statistics, and database systems. The overall goal of the data mining process is to extract information from a data set and transform it into an understandable structure for further use. Data Mining is widely used in diverse areas. There are number of commercial data mining system available today yet there are many challenges in this field.

2. Literature Review

Recently many Data mining research were done in the various domains, such as Mobile commerce. Paper [1] proposed cluster-Based Temporal Mobile Sequential pattern mine (CTMSP-Mine) to discover the cluster-Based Temporal Mobile sequential pattern. They used techniques two techniques such as Co-smart-cast algorithm to cluster the mobile transaction sequences. In this algorithm, they proposed LBS-alignment to evaluate similarity of mobile transaction sequences and GA-based time segmentation algorithm to find the most suitable time intervals. After clustering and segmentation user cluster table and time interval table are generated.CTMSP-Mine algorithm to mine CTMSPs from mobile transaction database according to user cluster table and time table. In online, they predict subsequent behaviors according to user's previous mobile transaction sequences and current time mining an interdisciplinary subfield of computer science is the computational process of discovering patterns in large data

sets involving methods at the intersection of artificial intelligence, machine learning, statistics, and database systems. The overall goal of the data mining process is to extract information from a data set and transform it into an understandable structure for further use. Data Mining is widely used in diverse areas. There are number of commercial data mining system available today yet there are many challenges in this field.

Paper [2] deals Students Mood recognition during online self-assessment test .They used exponential logic and its formulas for computation. Student's previous answers and slide bar status are considered as input. Total Number of questions for online self-assessment test, Student's goal, and slide bar value are used as variables for exponential logic. This system identifies student's current status of mood and gives appropriate feedback. Limitation of this system is student's manually selecting their mood using slide bar without any automation.

Paper [3] focused on how to improve aspect- level opinion mining for online customer reviews. They proposed the Joint Aspect/Sentiment model (JAS) to extract aspects and aspect-dependent sentiment lexicons from online customer reviews in a unified framework. They used Gibbs Sampling algorithm.

In Paper [4] a novel weakly supervised cybercriminal network mining method which can uncover both explicit and implicit relationships among cybercriminals based on their conversational messages posted on online social media. Mined two types of semantics such as transactional and collaborative relationships among cybercriminals using context-sensitive Gibbs sampling algorithm. They used

probabilistic generative model to extract multi-word expressions describing two types of cyber-criminal relationships in unlabeled messages. They used concept level approaches to better grasp the implicit semantics associated with text.

Research[5] deals about CIoT is a new network paradigm, where (physical/virutal) things or objects are interconnected and behave as agents with minimum human intervention, the things interact with each other following a context-aware perception-action cycle, use the methodology of understanding – by-building to learn from both the physical environment and social networks, store the learned semantic and/or knowledge in kinds of databases and adapt themselves to changes or uncertainties via resource-efficient decision-making mechanism. This research used game models and multiagent learning algorithm and these should be carefully designed for different applications in large-scale CIoT applications. This system (CIoT) needs massive sensitive data. Advantages of incorporation CIoT in applications are saving people's time and effort, Increasing resource efficiency and Enhancing service provisioning.

Paper [6] research on discovering and connections between social emotions and online documents as social affective text mining, including predicting emotions from online documents associating emotions with latent topics for document categorization to help online users to select related documents based on their emotional preferences. In this research, associate emotions with a specific emotional event/topic used instead of only a single term. They proposed a joint emotion-topic model for social affective text mining, which introduces an additional layer of emotion modeling into Latent Dirichlet Allocation (LDA). This model takes social affective text as input ex. College student jumps, affection related problem and categorize the text according to different emotions ex. Empathy, touched and surprise. They developed an approximate inference method based on Gibbs Sampling Algorithm.

Research [7] focused on classifying internet users based on their internet user behavior to offer him/her enhanced services. To do this, they have collected data such as Timestamp, IP address, URL and 10 keywords from proxy Cache and origin web server and categorized user behavior. To cluster users, two kinds of categorization algorithms are used such as "hard clustering" producing a partition, and an algorithm of "soft clustering" discovering overlapping clusters. The first algorithm is a method of hierarchical agglomerative clustering (HAC).

Research [8] on mobile social network services i.e., social networks and mobile devices applied them to social network groups with children with developmental disabilities. A foundation, university, hospitals and families with CDD created an online social network to share important information about services such as childcare. To access MSNSM users can use PDA, mobile devices and personal computers. These devices must have internet application or a web browser installed.

Research [9] proposed a real-time map generation and updating algorithm for car navigation system (CNS), which fulfills or corrects the unknown road area of the conventional road map database by using global positioning system information and wireless communication technologies. Here, introduced general algorithms for the extraction of road position, updating of road information, and updating cycle of the map database. But, updating and distribution of the road map database is a complex and high cost process.

Research [10] proposed frame work for Topic-Sensitive Influencer Mining (TSIM).In which, influence measurement of users and images is determined with hypergraph learning approach. It utilizes visual–textual content relations to construct homogeneous hyper edges for the topic distribution learning and social link relations to construct heterogeneous hyper edges for influence ranking in the network. They used hyper ranking algorithm, in which extracting topic-specific influential nodes such as content of images and social links in the networks such as flicker to identify the influence of users and scene or images.

Paper [11] represented text-categorization, in which a novel values assigned to a word are called distributional features, which include compactness of the appearance of the word and the position of the first appearance of the word. In this research, tfidf-style equation is constructed and ensemble learning technique is used. It considers frequency of a word i.e., how many times a word exists in a document and where the word appears at the very first time and compactness of words. This research work helps to improve performance while requiring only a little additional cost. In this research, just combines existing frequency with the distributional features.

Research [12] makes use of unsupervised network anomaly detection scheme based on TCM-KNN (Transductive Confidence Machines for k-Nearest Neighbors) data mining algorithm. This work address the optimizations for TCM-KNN in 2 aspects such as Genetic algorithm based instance selection method to limit the scale of training dataset and select most qualified instance to ensure the quality of dataset for training and Filter based feature selection method to extract the most necessary and relevant features to form the training set for TCM-KNN. These two optimization task are used to boost performance and reduce computational cost. This system found DDoS attacks, anomalies and take counter measurements in real network environment.

Research [13] It can help service users to locate the right service from the large number of available web services. It is computer as an aggregation of user's feedback rating over a specific period of time and reflects the reliability, trustworthiness and credibility of web service and its provider. This research makes use of Pearson correlation coefficient and Bloom filtering. It also recommends users when two or more web services have the same functionality but different Quality-of-Service (QoS) performance. This research used to increase the success ratio of composite services.

Research paper [14] introduces the concept of "data prospecting" to address the challenges of data intensive science. Research focuses on enhancing data selection through the use of interactive discovery engines. It enables researchers to filter the data based on the first look analysis,

discover interesting and previously unknown patterns to start new science investigations, verify the quality of data and corroborate whether the patterns in data match existing science theories or mental models. This research deals about data prospecting to earth science researchers. It provides discovery engine such as polaris Which uses the formula such as "more data usually beats better algorithm". The project has developed a mining algorithm to detect gap wind and associated ocean upwelling events globally. This research provides low cost technology.

In this Research paper [15], a novel approach for autonomous decision making is developed based on rough set theory of data mining. Algorithms which are used in this research are based on rough set theory, cluster analysis and measure theory. The two independent algorithms developed in this paper such as primary decision-making algorithm and confirmation algorithm to either generate an accurate diagnosis or make no decision. The proposed approach is illustrated with a medical case study involving diagnosis of patients with solitary pulmonary nodule using information from non-invasive tests. A nodule is detected on a patient's chest radiograph. The diagnosis is perceived depend on many features such as SPN diameter, border character, presence of calcification, patient's age, smoking history etc. It takes reduced number of features so it consumes lower testing costs.

The system proposed in Research paper [16] uses vast storage of information from medical data so that diagnosis based on these historical data can be made. It uses a unique algorithm which increases accuracy of such diagnosis by combining the key points of neural networks, Large Memory Storage, and Retrieval, k-NN, and differential diagnosis all integrated into one single algorithm. This algorithm can be used in solving diagnosis of multiple diseases showing similar symptoms, diagnosis of a person suffering from multiple diseases, receiving faster and more accurate second opinion, and faster identification of trends present in the medical records. This method is enhanced to reduce the huge number of underlying variables to just one by finding the root disease, or the most probable disease, using smart pattern matching involving k-NN classification technique and the next probable diseases by performing differential diagnosis, using the Hopfield neural networks theory and Large Memory Storage and Retrieval (LAMSTAR) Networks. This research work helps for saving time and money both the doctor and patient.

Research paper [17] proposed an algorithm for data mining called Ant-Miner (ant-colony-based data miner). It is an algorithm for clustering. The goal of Ant-Miner is to extract classification rules from data. It considers each case (ex., object, record or instance) and values of the attributes of the case (ex., gender=female) as input. It assigns each case to one class (ex.,if <conditions> then <class>. We compare the performance of Ant-Miner with CN2, a well-known data mining algorithm for classification, in six public domain data sets. ACO algorithm has the ability to perform a flexible robust search for a good combination of terms (logical conditions) involving values of the predictor attributes. Compare with CN2 it gives predictive accuracy.

Research paper [18] proposed a novel pattern mining approach to recognize sequential (brushing-washing face-preparing meals-eating meals) , interleaved(While eating meals going to bed room and answering for phone call and come back to eat),and concurrent (While eating watching TV) activities in a unified framework. Observations are collected by using physical sensors. Observations are used to train an appropriate activity model. The trained model can be used to assign new observations with activity labels. They proposed sliding-window-based algorithm which is commonly used in time series data analysis, trace segmentation algorithm to segment the trace accurately and epSICAR Activity Recognition Algorithm. A typical application of this work is monitoring Activities of Daily Living (ADLs) for the elderly and cognitively impaired people, and providing them with proactive assistance. In this research work does not include audio features. Due to finance and resource constraints data collected only in a real home.

Research paper [19] This paper proposed an online oversampling principal component analysis (osPCA) algorithm with Loo strategy to address the problem of most anomaly detection methods cannot be easily extended to large-scale problems without sacrificing computation and memory requirements, and we aim at detecting the presence of outliers from a large amount of data via an online updating technique. Our approach is especially of interest in online or large-scale problems which have computation or memory limitations. It gives both accuracy and efficiency.

In Research paper [20] proposed a technique called the user-expectation method. Given these expectations, the system uses a fuzzy matching technique to match the discovered patterns against the user's expectations, and then rank the discovered patterns according to the matching results. The proposed technique is general and interactive. This method considers subjective interestingness i.e., unexpectedness. This system finds the patterns from the database and asked the user expected patterns then executing matching algorithm with the use of fuzzy sets then ranks the discovered patterns. Limitations of this work are they did not claim that the problems associated with the interestingness or even the unexpectedness is fully understood. Further research is needed.

Research work [21] presented a novel method named structured cosine similarity (SCS) that furnishes speech document clustering with a new way of modeling on document summarization, considering the structure of the documents so as to improve the performance of document clustering in terms of quality, stability, and efficiency. When addressing document clustering, simple tfidf-based summarization spherical K-means clustering (SKC) was proven to outperform LSI in terms of the same level of clustering quality exhibited as LSI but with less time and memory required. In this research work, (e.x., customer service conversation) important words, numbers ex., credit card number from speech documents are retrieved (ignoring noises) and based on these data classification will be performed.

Research paper [22] introduces concepts and algorithms of

feature selection, surveys existing feature selection algorithms for classification and clustering, groups and compares different algorithms with a categorizing framework based on search strategies, evaluation criteria and data mining tasks, reveals unattempt combinations and provides guidelines in selecting feature selection algorithms. They developed categorizing framework to categorize different algorithms and give guide for intelligent feature selection algorithms and Unifying framework to create an integrated system that will automatically recommend the most suitable algorithm to user. To help users employ a suitable algorithm without knowing details of each algorithm.

Research paper [23] proposes a new inductive algorithm for image annotation by integrating label correlation mining and visual similarity mining into a joint framework. Automatic image annotation technology associates images with labels or tags. It enables conversion of image retrieval into text matching. Indexing and retrieval of text documents are faster and usually more accurate than that of raw multimedia data. It has high efficiency and accuracy. They proposed a new framework for automatic web and personal image labeling by integrating shared structure learning (SSL) and graph-based learning into a joint framework. The proposed framework, i.e., label correlation mining with relaxed graph embedding (LMGE).

Research work [24] deals about capture intention of camcorder users. Capture intention are two kinds of features: attention-specific features and content-generic features. Therefore, these two kinds of features are excerpted and analyzed based on video structure decomposition. Then singular value decomposition (SVD)-based intention segmentation and learning-based classification are performed based on these features.

Research [25] proposed a technique that uses partial information about the contents of a shopping cart for the prediction of what else the customer is likely to buy. Using the recently proposed data structure of item set trees (IT-trees), they obtain, in a computationally efficient manner, all rules whose antecedents contain at least one item from the incomplete shopping cart. Then, they combined these rules by uncertainty processing techniques, including the classical Bayesian decision theory and a new algorithm based on the Dempster-Shafer (DS) theory of evidence combination. The proposed algorithm speeds up the computation. The algorithm takes an incoming item set as the input and returns a graph that defines the association rules entailed by the given incoming item set. Proposed algorithms are depth-first search in the IT-tree and Simplified algorithm to update the rule graph. Experiments show that the computational costs of DS-ARM still grow very fast with the average length of the transactions and with the number of distinct items—in real world applications, this can become a serious issue.

In Research Paper [26] developed compact representation called "tiny videos" that achieves high video compression rates while retaining the overall visual appearance of the video as it varies over time. An exemplar-based clustering algorithm—achieves the best trade-off between compression and video recall. They used this large collection of user-labeled videos in conjunction with simple data mining

techniques to perform related video retrieval, as well as classification of images and video frames. The classification performance of the tiny videos data set was compared with tiny images. They showed that tiny videos are better suited for classifying sports-related activities than tiny images, while tiny images performed better at categorizing people. They showed that additional metadata in the tiny videos database can be used to significantly improve classification precision for some categories. Tiny video can be used for content-copy detection.

Research paper [27] deals issue in cost sensitive learning that considers both test costs and misclassification costs. This research work is to prove that "missing is useful" as missing values actually reduces the total cost of tests and misclassifications. They discussed and compared several strategies that utilize only known values and that "missing is useful" for cost reduction in cost-sensitive decision tree learning. They performed empirical experiments to compare the four strategies such as Known Value Strategy, Null Strategy, Internal Node Strategy and C4.5 Strategy on real-world data sets (medical data) by the total cost and concluded that the Internet Node Strategy is the best.

Research work [28] suggests new classification techniques for multichannel EEG recordings. Two time-series classification techniques, namely, Support Feature Machine (SFM) and Network-Based Support Vector Machine (SVM) (NSVM), are proposed in this paper to predict from EEG readings whether a person is epileptic or non-epileptic. The proposed SFM and NSVM techniques provide very promising and practical results and require much less time and memory resources than traditional techniques. The proposed techniques are tested on two EEG data sets acquired from ten and five patients, respectively.

Research [29] addresses the problem of router mis-configurations using data mining. They applied association rules mining to the configuration files of routers across an administrative domain to discover local, network-specific policies. They focused on three aspects of the configurations: user accounts, interfaces and BGP sessions. They used the approach called Minerals, applies data mining on router configuration files across a network to infer local, network-specific policies and detect potential errors that deviate from the inferred policies.

Research work [30] proposed an evolutionary game theoretic framework to model the dynamic information diffusion process in social networks. It helps to prevent the detrimental information spreading, e.g. computer virus. The influence of users' decisions, actions and socio-economic connections on information forwarding also plays an important role in the diffusion process. The social network topology can be treated as a graph structure and the user with new information can be regarded as the mutant. By considering the information diffusion process as the mutant spreading process, the graphical evolutionary game provides us with an ideal tool to find the final information diffusion state.

Research paper [31] focused on personalized travel recommendation by considering specific user profiles or attributes (e.g., gender, age, race) as well as travel group

types (e.g., family, friends, couple) automatically retrieved from freely available community-contributed photos. A probabilistic Bayesian learning framework which further entails mobile recommendation on the spot is introduced as well. This research did experiment on more than 10 million photos collected from 19 major cities worldwide. The proposed framework is promising to operate on mobile devices where the user profile and context information (e.g., geo-locations) can be easily detected from mobile sensors, and to further entail "location-awareness"—recommending next travel location from his/her current location or even delivering context-related advertisements or services.

Research work [32] forward a novel optimization algorithm based on the generative topic model. This work utilizes the semantic and temporal correlation among sequences and to build up a mutual reinforcement process. The proposed algorithm's performance is robust and stable.

Research paper [33] proposed a novel intelligent system which is able to automatically detect road accidents, notify them through vehicular networks, and estimate their severity based on the concept of data mining and knowledge inference. This system considers the most relevant variables that can characterize the severity of the accidents (variables such as the vehicle speed, the type of vehicles involved, the impact speed, and the status of the airbag). They developed a prototype based on off-the-shelf devices and validate it at the Applus + IDIADA Automotive Research Corporation facilities, showing that this system can notably reduce the time needed to alert and deploy emergency services after an accident takes place. We selected three of the classification algorithms such as Decision trees, Bayesian networks and Support Vector Machines provided by Weak to study which one obtains the best results. They finally concluded Bayesian models for classification.

Research paper [34] extends the ideas developed for feature selection problems to support classifier ensemble reduction, by transforming ensemble predictions into training samples, and treating classifiers as features. The target of classifier ensemble reduction (CER) is to reduce the amount of redundancy in a reconstructed classifier ensemble, to form a much reduced subset of classifiers that can still deliver the same classification results. CER is proposed which builds upon the ideas from existing FS techniques. This system provides superior classification performance.

Research paper [35] proposed a new research problem on active learning from data streams. The objective of this research is to label a small portion of stream data from which a model is derived to predict future instances as accurately as possible. To tackle the technical challenges raised by the dynamic nature of the stream data, i.e., increasing data volumes and evolving decision concepts, we propose a classifier ensemble-based active learning framework that selectively labels instances from data streams to build a classifier ensemble. They introduced a minimum-variance (MV) principle to guide the instance labeling process for data streams.

Research paper [36] Smartphone based web browsing wastes a lot of power when downloading web pages due to the special characteristics of the wireless radio interface. In this paper, we identify these special characteristics, and address power consumption issues through two novel techniques. We reorganize the computation sequence of the web browser when loading a webpage. This research work did experiment on Android phones on T-Mobile UMTS network. Experimental results proved that it can reduce the power consumption of smart phone by more than 30% during web browsing and also it reduces the webpage loading time and increase the network capacity. They proposed a low overhead prediction algorithm based on Gradient Boosted Regression Trees (GBRT).

Research [37] work proposed a novel computer vision based fall detection system for monitoring an elderly person in a home care, assistive living application. It helps, when an elderly person falls, a fall detection system will detect this fall event and an alarm signal will be sent to certain caregivers. Author's proposed a person-specific fall detection system based on a novel unsupervised algorithm which is termed as the online one class support vector machine (OCSVM) classifier. This Algorithm extracts three types of features such as ellipse features, shape structure features, and position features.

Research paper [38], investigates the identification of different variants for the process is explicitly accounted and based on the clustering of log traces. As finding an exact solution to such an enhanced process mining problem is proven to require high computational costs, in most practical cases, a greedy approach is devised. This paper proposed Mine workflow algorithm.

In this research paper [39], they proposed a novel framework, called Mobile Commerce Explorer (MCE), for mining and prediction of mobile users' movements and purchase transactions under the context of mobile commerce. The MCE framework consists of three major components: 1) Similarity Inference Model (SIM) for measuring the similarities among stores and items, which are two basic mobile commerce entities considered in this paper; 2) Personal Mobile Commerce Pattern Mine (PMCP-Mine) algorithm for efficient discovery of mobile users' Personal Mobile Commerce Patterns (PMCPs) and 3) Mobile Commerce Behavior Predictor (MCBP) for prediction of possible mobile user behaviors. This is the first work that facilitates mining and prediction of mobile users' commerce behaviors in order to recommend stores and items previously unknown to a user.

Research paper [40] addresses the problems of (a) crowdsourcing the annotation of first impressions of video bloggers (vloggers) personal and social traits in conversational YouTube videos, and (b) mining the impressions with the goal of modeling the interplay of different vlogger facets. Their study includes 442 YouTube vlogs and 2,210 annotations collected in Mechanical Turk. The algorithm used Latent Dirichlet Allocation (LDA) model to discover joint first impressions of vlogger personality, attractiveness and mood.

3. Research trend in Data Mining

3.1 Data collection

In this survey, we have collected research papers published in the field of Data mining for past 30 years through internet. Data samples picked were 40 IEEE research papers. Out of which, predominantly 37 papers were published in timeline 2004-2014 year, 3 papers lies between 1993-2003 and no papers in between 1982-1992.

The above details were depicted in the below tables and plotted in graphs.

Table 1: Data Mining Research Paper publication

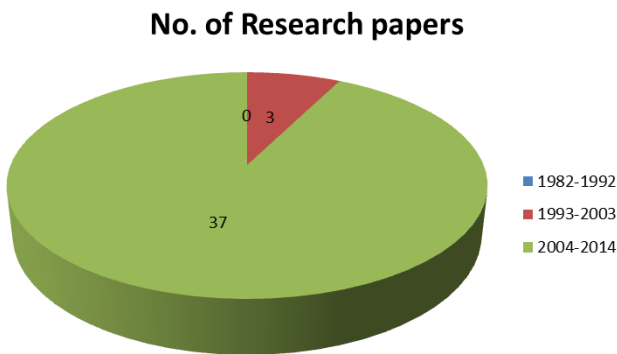| Years	No. of Research papers
1982-1992	0
1993-2003	3
2004-2014	37

Figure 1: Year wise comparison

4. Research and its Applications

The Main focus is that we are categorizing the 40 research papers based on their application areas.

Paper [1] and [39] contribution for Mobile commerce,[2] Students Mood Analyses, Paper[3],[5],[6],[7],[10],[11],[13], [21],[23],[24],[26],[30],[31],[32],[36] and [40] Researches for an efficient Information suggestion and retrieval, Research paper [4],[12] and [19] were focused on Intrusion detection, Research Paper[8],[18] and [37] were focused for Helping disabled and elder people, Paper[9] and [33] research for an Automobiles applications, Research paper[14],[15],[17],[20],[22],[34],[35] and [38] focuses the Advancements in Data Mining techniques, Paper[16] and [28] research work used for Analyzing Medical data, Paper [25] and [27] focus were on to analyze Missing items, Paper[29] Research for Misconfiguration of Routers in Wireless applications.

Table 2: Data Mining –Applications

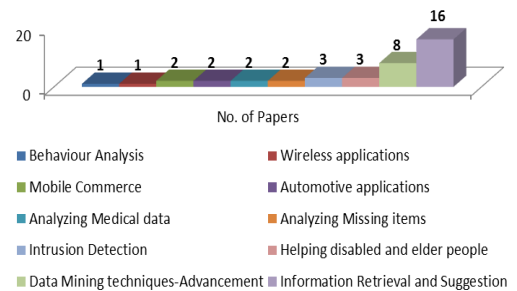| Application areas	No. of Papers
Mobile Commerce	2
Behavior Analysis	1
Information Retrieval and Suggestion	16
Intrusion Detection	3
Helping disabled and elder people	3
Automotive applications	2
DataMining Advancement techniques	8
Analyzing Medical data	2
Analyzing Missing items	2
Wireless applications	1

Figure 2: Applications trend Pareto wise comparison

5. Conclusion

Data mining or Knowledge Data Discovery is the computer-assisted process of digging through and analyzing enormous sets of data and then extracting the meaning of the data. It is applied effectively not only in business environment but also in other fields such as weather forecast, medicine, transportation, healthcare, insurance, government…etc.

We have discussed about each paper elaborately and categorized these papers based on their application areas. In this survey paper, we have analyzed 40 research papers from different application fields of Data Mining. From this study, we could conclude that data mining is an ever growing research field with interdisciplinary applications increasing.

6. Acknowledgement

References

[1] Eric Hsueh-Chan Lu, Vincent S. Tseng, Member, IEEE, and Philip S. Yu, Fellow, IEEE "Mining Cluster-Based Temporal Mobile Sequential Patterns i n Location-Based Service Environments" IEEE TRANSAC TIONS ON KNOWLEDGE AND DATA ENGINE ERING, VOL. 23, NO. 6, JUNE 2011

[2] Christos N. Moridis and Anastasios A. Economides "Mood Recognition during Online Self-Assessment Tests" IEEE TRANSACTIONS ON LE ARNING TECHN OLOGIES, VOL. 2, NO. 1, JANUARY-MARCH 2009

[3] XU Xueke, CHENG Xueqi, TAN Songbo, LIU Yue, SHEN Huawei "Aspect-Level Opinion Mining of Online Customer Reviews " Key Laboratory of Web Data Science and Technology, Institute of Computing Technology, Chinese Academy of Sciences, Beijing 100190, China, MANAGEMENT AND VISUALIZATION OF USER AND NETWORK DATA, China Communications, March 2013

[4] Raymond Y.K. Lau Department of Information Systems, City University of Hong Kong, Hong Kong SAR Yunqing Xia Department of Computer Science and Technology, Tsinghua University, Beijing 100084, CHINA Yunming Ye Shenzhen Key Laboratory of Internet Information Collaboration, Shenzhen graduate School, Harbin Institute of Technology, Shenzhen 518055, CHINA "A Probabilistic Generative Model for Mining Cybercriminal Networks from Online Social Media"

[5] Qihui Wu, Senior Member , IEEE, Guoru Ding, Student Member, IEEE , Yuhua Xu, Student Member , IEEE, Shuo Feng, Zhiyong Du, Jinlong Wang, Senior Member , IEEE, and Keping Long, Senior Member , IEEE "Cognitive Internet of Things: A New Paradigm Beyond Connection" IEEE INTERNET OF THINGS JOURNAL, VOL. 1, NO. 2, APRIL 2014

[6] Shenghua Bao, Shengliang Xu, Li Zhang, Rong Yan, Zhong Su, Dingyi Han, and Yong Yu "Mining Social Emotions from Affective Text "IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, VOL. 24, NO. 9, SEPTEMBER 2012

[7] Luigi Lancieri, Member, IEEE, and Nicolas Durand "Internet User Behavior: Compared Study of the Access Traces and Application to the Discovery of Communities" IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS—PART A: SYSTEMS AND HUMANS, VOL. 36, NO. 1, JANUARY 2006

[8] Li-Der Chou, Member, IEEE, Nien-Hwa Lai, Yen-Wen Chen, Member, IEEE, Yao-Jen Chang, Jyun-Yan Yang,Lien-Fu Huang, Wen-Ling Chiang, Hung-Yi Chiu, and Haw-Yun Shin "Mobile Social Network Services for Families With Children With Developmental Disabilities" IEEE TRANSACTIONS ON INFORMATION TECHNOLOGY IN BIOMEDICINE, VOL. 15, NO. 4, JULY 2011

[9] SeungGwan Lee, DaeHo Lee, and SungWon Lee, Member, IEEE "Network-Oriented Road Map Generation for Unknown Roads using Visual Images and GPS-based Location Information" S.G. Lee et al.: Network-Oriented Road Map Generation for Unknown Roads using Visual Images and GPS-based Location Information

[10] Quan Fang, Jitao Sang, Changsheng Xu, Fellow, IEEE, and Yong Rui, Fellow, IEEE," Topic-Sensitive Influencer Mining in Interest-Based Social Media Networks via Hypergraph Learning", IEEE TRANSACTIONS ON MULTIMEDIA, VOL. 16, NO. 3, APRIL 2014

[11] Xiao-Bing Xue and Zhi-Hua Zhou, Senior Member, IEEE, "Distributional Features for Text Categorization", IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, VOL. 21, NO. 3, MARCH 2009

[12] Yang Li Institute of Computing Technology, Chinese Academy of Sciences, PH.D., Kexueyuan South Rd. 6, Beijing China, Bin-Xing Fang Institute of Computing Technology, Chinese Academy of Sciences,Professor, Kexueyuan South Rd. 6, Beijing China

[13] Shangguang Wang, Member, IEEE, Zibin Zheng, Member, IEEE, Zhengping Wu, Member, IEEE, Fangchun Yang, Member, IEEE, Michael R. Lyu, Fellow, IEEE, "Reputation Measurement and Malicious Feedback Rating Prevention in Web Service Recommendation Systems", IEEE TRANSACTIONS ON SERVICES COMPUTING, VOL. , NO. , MARCH 2014.

[14] Rahul Ramachandran, John Rushing, Amy Lin, Helen Conover, Xiang Li, Sara Graves, U. S. Nair, Kwo-Sen Kuo, Member, IEEE, and Deborah K. Smith," Data Prospecting–A Step Towards Data Intensive Science", IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING, VOL. 6, NO. 3, JUNE 2013.

[15] Andrew Kusiak, Jeffrey A. Kern, Kemp H. Kernstine, and Bill T. L. Tseng, "Autonomous Decision-Making: A Data Mining Approach", IEEE TRANSACTIONS ON INFORMATION TECHNOLOGY IN BIOMEDICINE, VOL. 4, NO. 4, DECEMBER 2000.

[16] Rahul Isola, Student Member, IEEE, Rebeck Carvalho, Student Member, IEEE and Amiya Kumar Tripathy, Member, IEEE," Knowledge Discovery in Medical Systems Using Differential Diagnosis, LAMSTAR, and k-NN ", IEEE TRANSACTIONS ON INFORMATION TECHNOLOGY IN BIOMEDICINE, VOL. 16, NO. 6, NOVEMBER 2012.

[17] Rafael S. Parpinelli, Heitor S. Lopes, Member, IEEE, and Alex A. Freitas," Data Mining With an Ant Colony Optimization Algorithm", IEEE TRANSACTIONS ON EVOLUTIONARY COMPUTING, VOL. 6, NO. 4, AUGUST 2002.

[18] Tao Gu, Member, IEEE, Liang Wang, Student Member, IEEE, Zhanqing Wu, Xianping Tao, Member, IEEE, and Jian Lu," A Pattern Mining Approach to Sensor-Based Human Activity Recognition", IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, VOL. 23, NO. 9, SEPTEMBER 2011.

[19] Yuh-Jye Lee, Yi-Ren Yeh, and Yu-Chiang Frank Wang, Member, IEEE, "Anomaly Detection via Online Oversampling Principal Component Analysis Yuh-Jye", IEEE TRANSACTIONS ON

KNOWLEDGE AND DATA ENGINEERING, VOL. 25, NO. 7, JULY 2013.

[20] Bing Liu, Member, IEEE Computer Society, Wynne "I,Lai-Fun Mun, and Hing-Yan Lee, Member, IEEE Computer Society,'" Finding Interesting PatternsUsing User Expectations", IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, VOL. 11, NO. 6, NOVEMBERIDECEMBER 1999.

[21] Soe-Tsyr Yuan and Jerry Sun," Ontology-Based Structured Cosine Similarity in Document Summarization: With Applications to Mobile Audio-Based Knowledge Management", IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS—PART B: CYBERNETICS, VOL. 35, NO. 5, OCTOBER 2005.

[22] Huan Liu, Senior Member, IEEE, and Lei Yu, Student Member, IEEE, "Toward Integrating Feature Selection Algorithms for Classification and Clustering", IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, VOL. 17, NO. 4, APRIL 2005.

[23] Yi Yang, Fei Wu , Feiping Nie, Heng Ta o Shen, Yueting Zhuang, and Alexander G. Hauptmann," We b and Personal Image Annotation by Mining Label Correlation With Relaxed Visual Graph Embedding" , IEEE TRANSACTIONS ON IMAG E PROCESSING, VOL. 21, NO. 3, MARCH 2012.

[24] Tao Mei, Member, IEEE, Xian-Sheng Hua, Member, IEEE, He-Qin Zhou, and Shipeng Li, Member, IEEE," Modeling and Mining of Users' Capture Intention for Home Videos", IEEE TRANSACTIONS ON MULTIMEDIA, VOL. 9, NO. 1, JANUARY 2007.

[25] Kasun Wickramaratna, Student Member , IE EE , Miroslav Kubat, Senior Member , IEEE, and Kamal Premaratne, Senior Member , IEEE," Predicting Missing Items in Shopping Carts", IEEE TRANSAC TIONS ON KNOWLEDG E AND DATA ENGINEER ING, VOL. 21, NO. 7, JULY 2009.

[26] Alexandre Karpenko, Student Member, IEEE, and Parham Aarabi, Senior Member, IEEE," Tiny Videos: A Large Data Set for Nonparametric Video Retrieval and Frame Classification", IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, VOL. 33, NO. 3, MARCH 2011.

[27] Shichao Zhang, Senior Member, IEEE, Zhenxing Qin, Charles X. Ling, and Shengli Sheng," "Missing Is Useful": Missing Values inCost-Sensitive Decision Trees", IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, VOL. 17, NO. 12, DECEMBER 2005.

[28] Wanpracha Art Chaovalitwongse, Senior Member, IEEE, Rebecca S. Pottenger, Shouyi Wang, Ya-Ju Fan, and Leon D. Iasemidis, Member, IEEE," Pattern- and Network-Based Classification Techniques for Multichannel Medical Data Signals to Improve Brain Diagnosis", IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS—PART A: SYSTEMS AND HUMANS, VOL. 41, NO. 5, SEPTEMBER 2011.

[29] Franck Le, Sihyung Lee, Student Member, IEEE, Tina Wong, Hyong S. Kim, and Darrell Newcomb," Detecting Network-Wide and Router-Specific Misconfigurations Through Data Mining", IEEE/ACM TRANSACTIONS ON NETWORKING, VOL. 17, NO. 1, FEBRUARY 2009.

[30] Chunxiao Jiang, Member, IEEE, Yan Chen, Member, IEEE, and K. J. Ray Liu, Fellow, IEEE," Graphical Evolutionary Game for Information Diffusion Over Social Networks", IEEE JOURNAL OF SELECTED TOPICS IN SIGNAL PROCESSING, VOL. 8, NO. 4, AUGUST 2014.

[31] Yan-Ying Chen, An-Jung Cheng, and Winston H. Hsu, Senior Member, IEEE," Travel Recommendation by Mining People Attributes and Travel Group Types From Community-Contributed Photos", IEEE TRANSACTIONS ON MULTIMEDIA, VOL. 15, NO. 6, OCTOBER 2013

[32] Xiang Wang, Xiaoming Jin, Meng-En Chen, Kai Zhang, and Dou Shen," Topic Mining over Asynchronous Text Sequences", IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, VOL. 24, NO. 1, JANUARY 2012.

[33] Manuel Fogue, Piedad Garrido, Member, IEEE, Francisco J. Martinez, Member, IEEE, Juan-Carlos Cano, Carlos T. Calafate, and Pietro Manzoni, Member, IEEE," A System for Automatic Notification and Severity Estimation of Automotive Accidents", IEEE TRANSACTIONS ON MOBILE COMPUTING, VOL. 13, NO. 5, MAY 2014.

[34] Ren Diao, Fei Chao, Member, IEEE, Taoxin Peng, Neal Snooke, and Qiang Shen," Feature Selection Inspired Classifier Ensemble Reduction", IEEE TRANSACTIONS ON CYBERNETICS, VOL. 44, NO. 8, AUGUST 2014.

[35] Xingquan Zhu, Peng Zhang, Xiaodong Lin, and Yong Shi," Active Learning From Stream Data Using Optimal Weight Classifier Ensemble", IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS—PART B: CYBERNETICS, VOL. 40, NO. 6, DECEMBER 2010.

[36] Bo Zhao, Wenjie Hu, Qiang Zheng, and Guohong Cao," Energy-Aware Web Browsing on Smartphones", Citation information: DOI 10.1109/TPDS.2014.2312931, IEEE Transactions on Parallel and Distributed Systems.

[37] Miao Yu, Yuanzhang Yu, Adel Rhuma, Syed Mohsen Raza Naqvi, Member, IEEE, Liang Wang, Senior Member, IEEE, and Jonathon A. Chambers, Fellow, IEEE," An Online One Class Support Vector Machine-Based Person-Specific Fall Detection System for Monitoring an Elderly Individual in a Room Environment", IEEE JOURNAL OF BIOMEDICAL AND HEALTH INFORMATICS, VOL. 17, NO. 6, NOVEMBER 2013.

[38] Gianluigi Greco, Antonella Guzzo, Luigi Pontieri, and Domenico Sacca` , Member, IEEE Computer Society," Discovering Expressive Process Models by Clustering Log Traces", IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, VOL. 18, NO. 8, AUGUST 2006.

[39] Eric Hsueh-Chan Lu, Wang-Chien Lee, Member, IEEE, and Vincent S. Tseng, Member, IEEE," A Framework for Personal Mobile Commerce Pattern Mining and Prediction", IEEE TRANSACTIONS ON

KNOWLEDGE AND DATA ENGINEERING, VOL. 24, NO. 5, MAY 2012.

[40] Joan-Isaac Biel and Daniel Gatica-Perez, Member, IEEE," Mining Crowdsourced First Impressions in Online Social Video", Citation information: DOI 10.1109/TMM.2014.2346471, IEEE Transactions on Multimedia.

Author Profile

Mrs. Dhanalakshmi D has finished her Bachelor of Computer Science/Computer Applications from Avinashilingam Institute for Home Science and Higher Education for Women Coimbatore. She has completed her M.C.A from Bharathiar University. She has been awarded her M.Phil in Data Mining from Bharathiar University. She worked as Assistant professor in Dr.NGP Arts and Science College Coimbatore for four year. She is currently a Full time Research Scholar in Department of Computer Science at SNR SONS COLLEGE, Coimbatore, Tamil Nadu, India working towards her Ph.D.

Mrs J.Komala Lakshmi has finished her B.Sc Mathematics from SeethaLakshmi Achi College for women ,Pallathur, during 1995. She has completed her M.C.A from J.J.college Of Arts and Sciences, Pudhukottai during 1998. She has been Awarded her M.Phil Computer Science, Bharathiar University during 2008. She has been Awarded with the Silver medal for her academic Excellence in M.C.A from the Institution. She has ten years of teaching experience in Collegiate service. She is currently working as Assistant Professor, Department of Computer Science, SNR SONS COLLEGE, Coimbatore. She has presented her papers in three international conferences. She has published papers in six international journals .She also contributed a book publications to the computer science students. She was a Graduate Student Member of IEEE. She has been awarded with the recognition from Who's Who in the World 2011s, 28th Edition for her notable contribution to the society.