

Applications of Data Mining – A Survey

Lovedeep¹ and Sabia²

^{1,2}*Department of Computer Sc. & Engineering,
Guru Nanak Dev University Regional Campus, Jalandhar, India
E-mail: ¹Lovedeep1991@gmail.com, ²sabijal@gmail.com*

Abstract—Data mining is an analytic process designed to explore large amount of data and to predict new subset of data. There are different MNC's and organizations that operate at different places in different countries and generate massive data at each operation place. The decision making process requires to gather all data and with help of different data mining techniques the useful information is abstracted from large data items. In this paper, the main focus is how data mining is useful in different organizations such as retail stores, banks, hospitals and insurance companies where we have to deal with massive data and list the applications of data mining in different fields.

Keywords: *Data Mining Model, Decision Making Using Data Mining, Stages, Data Mining, KDD*

I. INTRODUCTION

Now a day people around the world are using vast data in different fields such as health care, business, industries etc. and this data is not in same format everywhere, we are having different forms of data such as documents, graphics, audio formats etc. Data mining is basically retrieving useful information from these huge databases. This technique is also called KDD (knowledge discovery process). We are having large amount of data but there is lack of useful information that is “we are data rich but information poor”. It is very hard process to turn large data volumes into useful and worth decision making knowledge. To solve this problem of retrieving useful information from large set of data, we are moving toward the field of data mining. To take maximum advantage of raw data only its retrieval is not enough. It require different tools and methods for automated submission of data, extraction of useful items and information from it, discovery and prediction of new patterns and interpretation of data for decision making process. This all can be done efficiently by “data mining”.

Data mining is to extract hidden predictive information from large databases and it helps organizations to focus them on most important information of their warehouses [1, 2]. Data mining tools are useful in predicting the future trends and behaviours which are helpful for organizations to take proactive knowledge driven decisions. It is also saves our time as data is reduced after information is exacted from it. The data mining process consist of three major stages are:

1. Initial exploration: Exploration means to search or investigate. In this stage, the data preparation is done which involve cleaning data, data transformations, selecting subsets of records. In case of having large number of variables or fields

in sub sets, we perform some preliminary feature operations to keep variables in a manageable range.

2. Model building and validation: In next stage the relevant variables and complexity is determined. Various models are considered bases on the predictive performance. It is very long process and various techniques are used to compare different models and chose best one called competitive evaluation of models.
3. Deployment: The model which is selected in the last stage is as best model is used and new data values are applied to it for future estimations and predictions.

This paper is organised as follows. In section II literature survey. In Section III the various applications has been discussed. Final section gives a conclusion of the paper.

II. LITRATURE SURVEY

1. Dominik Fisch [7] in 2014 author introduce new way of fusing classifiers at the level of parameters of classification rules the two novel technique to fuse probabilistic generative classifiers(CMM) into one. This technique based on multinomial distributions for categorical input and multivariate normal distributions. The main advantage of this fusion is to hyper distributions throughout the fusion process used components used in online training.
2. Jianlin Xu, [8] in 2013 to improve the security status of mobile app author propose a methodology for mobile apps based on data mining and cloud computing to filter out malware apps from mobile app markets and also present prototype system is mobsafe. Mobsafe combined the static and dynamic analysis methods are SAAF and ASEF to estimate the total time needed to evaluate all the apps stored in one mobile app market.
3. Neelamadhab Padhy, [16] in 2012The author talks about the two major applications of data mining i.e. generic applications and domain specific applications. It is observed that no generic application is fully generic. There are limitations of generic applications of data mining. Domain and data, context parameters and aim of data mining try to influence the data mining decisions. Domain specific applications produce more accurate results which are over 90% and these are more specific for data mining. It is difficult to design such mining system which works for any domain dynamically.

4. Shouyi Wang [4] in 2011 author describes that numerical errors are detectable in advance; data mining techniques can be used for early detection of these numerical typing errors. Author used multichannel electroencephalogram (EEG) recordings for quantitative analysis of detecting errors along with two basic data mining techniques i.e. linear discriminate analysis (LDA) and support vector machine (SVM). Around 80% errors can be detected even without processing data saving lot of time. Using data mining techniques it is possible to proactively predict the keystrokes with errors based on EEG recordings. The only drawback is that the study is based on limited data pool, it maybe not produce result in generalized form.
5. Mahdi Esmaeili [5] in 2010 author presented there is huge number of variables and objectives involved in aerospace engineering optimization which cannot be ignored, so only multi-objective optimization can deal with it effectively and also proposed method to use data mining techniques for aerospace engineering optimization process. The advantage of using these techniques is that less variables decrease the effective cost of optimization. Here simple variable reduction tool and some data mining techniques are applied for getting desired results. Three algorithms are used which speed up the optimization process. Even in worse case 55% variables are reduced and BFTree and J48 algorithms require fewer variables while LAD Tree algorithm utilizes at least 7 variables to classify data set.
6. Sérgio Ramos, Zita Vale [6] in 2008 author present an electricity medium voltage (MV) consumers characteristics and classifications and also compare the three different clustering algorithms for taking the number of clusters and it also presents the new tariff structures to apply for each customer class.
7. Bartley D. Richardson [9] in 2008 author describes the techniques used by fellow's reactions and feedback from the students by integrates data mining and software engineering. STEP project is based on engineering principles and relevant to student's lives. To increase the confidence and learning abilities of students using STEM subject's mega mining mart is designed to understand correlations in a familiar setting and periodic table.
8. Elovici, Y., Kandel [12] author proposed the knowledge based methodology to view terror-related context at a series of evasive web sites. This methodology use data mining algorithm to the textual context of terror-related web sites. User can view all activities done by terrorist. Author also describes the intrusion detection system (IDS), vector space model and clustering techniques.

III. APPLICATIONS

A. Medical Science and Health Care Field

There is constant work going on in field of healthcare and medical sector to improve the facilities for patients. At same time it is hard to deal with data captured around healthcare processes in form of electronic health records, health insurance claims, disease registries and clinical trials. Data getting huge day by day and task here is to not only store data but use it effectively so as to improve medical facilities. Data mining can be used for this purpose as medical data is very complex and difficult to analyze. The data mining algorithms are helpful in reducing patient's risks and diagnosis costs. Using the prediction algorithms the observed prediction accuracy was 100% for 91.3% cases [3]. Also the success in health care depends upon the availability of clean data. Data should be capture properly, stored in proper manner for better mining of data.

B. Market Analysis and Retail Stores

With market segmentation, it is easy to find out behaviours that are same among the customers. You can look for patterns among customers that seem to purchase the same products at the same time. Another application of data mining is called customer churn. Customer churn helps in estimating that which customers are likely to stop purchasing products or services from you and go to one of your competitors. Also any company can use data mining to find out which purchases are the most likely to be fraudulent. Using data mining a retail store may be able to find out the products which are stolen maximum number of time. By finding out which products are stolen the most, steps can be taken to protect those products and detect those who are stealing them. Sequential pattern mining is effective mining method used in field of marketing.

C. Predictions in Engineering Field

In recent years, there is increase in the use of data mining techniques on such artefacts are with the goal of analyze and improving software processes for a given organization or project. As data mining provides techniques like intentional mining for predictions of data, it can be used in field of engineering. We can use data mining for predictions of cost estimation in engineering field.

D. Network Security

Security within internet is very serious issue. The e-business culture is developing more and more in recent years as effect there is more threat posed by internet crime. Data mining techniques such as association rule and clustering analysis are quite useful in preventing the network from these threats. Data mining utilize the light weight statistical summaries

which are gathered at distributed points within a network for detecting the security threats with help of signature detection filtering mechanism [10].

E. The Intelligence Agencies

Data mining is quite useful in field of intelligent agencies. These agencies have to deal with terrorists and other threats to countries. With help of data mining, agencies collect data regarding the activities of terrorists to investigate their future attacks. The Clustering technique is helpful (Association rule mining) for the different objects (like persons, organizations, vehicles etc.) in crime records. Data mining detects as well as analyzes the crime data. Agencies are developing new algorithms which make it easy to work with large data set [12].

F. Data Mining in Field of Sports

The world of sports produces a huge statistical data about each team and player. This data can be very useful for the assist coaches and managers if it can be managed properly by experts. Now a day sports data mining is very popular trend for prediction of result and player assessment. Even it can be utilized for identification of new talent, predictions about injuries and game strategies. By using data mining tools the experts can help players to improve the performance. Data mining techniques such as ANN, decision trees, Bayesian method, SVM, logistic regression and fuzzy methods have been employed to predict game results. By using hybrid algorithms the accuracy of prediction can be increased [11].

G. Web Education Systems

The courseware can be improved by using data mining techniques in web education. Data can be picked from student sessions on web and can be utilized by teachers or authors to improve the content and effectiveness of course. Data mining help increasing awareness among the learners [13].

H. Broadcasting Data in Mobile Computer Scenario

By using wireless communication devices, mobile can be accessed from any place. As we know downlink speed is more as compared to uplink speed, we can use data mining techniques in broadcasting of data. There are two methods used for this purpose. We try to arrange the subsequently requested data items are places closed to each other. It is also seen that using data mining in broadcasting data, the access latency can be decreased very efficiently [14].

I. Financial and Investment Field

Economic growth is one of the major issues for developing countries like India. Information technology plays great role in economic growth and globalization of economy. Due to this vast financial data being

generated day by day and accumulated at unprecedented place. Data mining provides automated approach for utilization of vast data and to give back information to financial companies for better investment strategies. Hidden patterns are discovered by data mining techniques like Sequential pattern and time-series mining, clustering analysis and association rules [15].

IV. SUMMARY

TABLE I

Sr. No	Data Mining Methods and Techniques	Applications Field	Type of Data Set
1.	Clustering Technique	Bio-Informatics, Medical Sciences, Marketing and Business, World Wide Web	Statistical Data, Discrete and Comparative Numerical Data.
2.	Decision Tree	Operational Research, Field of Sports, Sales and Purchase	Graphical Data, Trees and Graphs
3.	Sequential Pattern Mining	Retail Marketing, Financial and Investment Field	Continuous Data, Numeric or Alphanumeric
4.	Intentional Mining	Web Search, Business, Engineering	Interactive Data from Device Such as Computers, Images etc
5.	Association Rule	Intelligence Agencies, Security of Networks	Large Variable Data Set, Numerical, Alphanumeric, Visual and Audio Data
6.	Factor Analysis	Field of Psychology, Intelligencer Verbal Intelligence	Statistical Data, Large and Numerical values Visual Data Sets

V. CONCLUSION

The paper provided various applications of data mining and the mining methods that are used in different fields. As we know that data varies such as text, graphical and audio data in different fields so it is important to deploy a method which reduce variable in reasonable way by using data mining tools and methods such as clustering and sequential pattern mining. Depending upon the data type available in a particular field, method is selected to make data mining process convenient and swift. Evolving data mining applications have shown great potentials in financial, medical, network security and engineering fields and will continue to prosper in new fields such as sports and web education. By using proper mining techniques more applications can be explored in future with higher efficiency.

REFERENCES

- [1] Larose, D.T., "Discovering Knowledge in Data: An Introduction to Data Mining" 0-471-66657-2, John Wiley & Sons, Inc, 2005.
- [2] Dunham, M.H., Sridhar S., "Data Mining: Introductory and Advanced Topics", Pearson Education 81-7758-785-4, 1st Edition, 2006.

- [3] Neelamadhab Padhy, Dr. Pragnyaban Mishra, Rasmita Panigrahi “ Survey of Data Mining Applications And Feature Scope” IJCSEIT, (Vol. 2) June 2012.
- [4] Shouyi Wang, Cheng-Jhe Lin, Changxu Wu, and Wanpracha Art Chaovaitwongse “Early Detection of Numerical Typing Errors Using Data Mining Techniques” IEEE November 2011.
- [5] Mahdi Esmaeili, Amirhosein Mosavi, “Variable Reduction for Multi Objective Optimization Using Data Mining Techniques; Application to Aerospace Structures” 2010 2nd International Conference on Computer Engineering and Technology.
- [6] Sérgio Ramos, Zita Vale “Data Mining techniques application in Power Distribution utilities” IEEE 2008.
- [7] Dominik Fisch, Edgar Kalkowski, and Bernhard Sick “Knowledge Fusion for Probabilistic Generative Classifiers with Data Mining Applications” IEEE Transactions (Vol. 26) 3, March 2014.
- [8] Jianlin Xu, Yifan Yu, Zhen Chen_, Bin Cao, Wenyu Dong, Yu Guo, and Junwei Cao “MobSafe: Cloud Computing Based Forensic Analysis for Massive Mobile Applications Using Data Mining”.
- [9] Tsinghua Science and Technology ISSN (Vol. 18), 4, August 2013, 418–427.
- [10] P.J. Sandford “Detecting security threats in network core using data mining techniques” ISSN-1-4244-0143-7/06/20 IEEE 2006.
- [11] Maral Haghighat, Hamid Rastegariand, Nasim Nourafza “A Review of Data Mining Techniques for Result Prediction in Sports” ACSIJ Advances in Computer Science: an International Journal, Vol. 2, Issue 5, No. 6, November 2013. ISSN: 2322–5157.
- [12] Elovici, Y., Kandel, A., Last, M., Shapira, B., Zaafrany, O., “Using Data Mining Techniques for Detecting Terror-Related Activities on the Web”.
- [13] Romero, C., Ventura, S. and De-Bra, P. “Knowledge Discovery with Genetic Programming” for Providing Feedback to Courseware Authors, Kluwer Academic Publishers, Printed in the Netherlands, 2004”.
- [14] Dongsong Zhang and Lina Zhou ” Discovering Golden Nuggets: Data Mining in Financial Application” IEEE Transactions, Vol. 34, 2004.
- [15] Dongsong Zhang and Lina Zhou ” Discovering Golden Nuggets: Data Mining in Financial Application” IEEE Transactions Vol. 34, 2004.
- [16] Neelamadhab Padhy, Dr. Pragnyaban Mishra, Rasmita Panigrahi “ Survey of Data Mining Applications And Feature Scope” (IJCSEIT), Vol. 2, No. 3, June 2012.