

Review of Data Mining Tools, Techniques and Applications

Prachi Deshmukh-Chaudhari,
Bangalore, India.

Abhijeet Deshmukh,
Bangalore, India.

Abstract: Data mining is a rapidly growing field which has wide applications in variety of fields. It is a multi-disciplinary field which integrates statistics, neural networks, machine learning, visualization etc. This paper is an attempt to briefly review the various tools and techniques used in data mining. The paper also reviews some of the important applications of data mining in various areas.

Keywords: data Mining; tools, techniques; applications.

INTRODUCTION:

Data mining is an interdisciplinary field. It is the process that results in the discovery of new patterns in large data sets. It utilizes methods which are derived from artificial intelligence, machine learning, statistics, database systems etc. The paper is divided into three sections in which the important tools, techniques and applications of data mining are reviewed.

1. DATA MINING TOOLS:

Most of the data mining tools can be classified into three major categories:

1. Traditional data mining tools

2. Dashboards

3. Text-mining tools

1.1 Traditional data mining tools:

Traditional data mining tools such as OLAP (On Line Analytical Processing) help companies to establish data patterns and trends by using a number of complex algorithms and techniques.

1.2 Dashboards:

Dashboards are installed in computers to monitor information in a database. They reflect data changes and updates onscreen. It can be a chart, graph or table which

enables the user to observe

the business performance. Dashboards are easy to use and preferred because they can give an overview of the company's performance at a glance.

1.3 Text-mining tools:

These tools scan content of a document and convert the selected data into a format that is compatible with the tool's database, thus providing users with an easy and convenient way of accessing data without the need to open different applications. Some of the examples of text mining softwares are RapidMiner, FairIsaac, OpenNLP, IBM LanguageWare, SAS etc.

2. DATA MINING TECHNIQUES:

Data mining is an inter-disciplinary field which involves Statistics, Machine Learning, neural networks, visualization, etc. Hence it has adopted many techniques from these fields.

The most commonly used techniques include statistical approach, machine learning approach, artificial neural networks, visualization etc. Each of these techniques analyzes data in different ways.

2.1 Statistical Approach:

In this approach statistical models such as Bayesian network, regression Analysis, correlation analysis, cluster analysis are used. Statistical models are built on a set of data. Based on predefined measures the statistical model derives patterns, rules.

2.2 Artificial Neural Networks (ANN):

ANNs are non-linear predictive models. They are powerful predictive modeling techniques. ANNs can be used while reviewing records to identify fraud and fraud-like actions. As ANNs are complex to implement, they are used in situations where they can be reused for

multiple times, such as reviewing credit card transactions every month to check for anomalies.

2.3 Machine Learning Approach:

Decision tree algorithm, Inductive concept learning, conceptual clustering are some of the machine learning approaches used for data mining. Machine Learning methods search for the best model matching the data set.

2.4 Database based approach:

Data models or data base specific heuristics are used to derive the characteristics of the dataset.

Attribute oriented induction, attribute focusing, iterative database scanning for frequent data sets are some of the database based approach.

3. DATA MINING APPLICATIONS:

In this section some of the important applications of Data Mining are discussed.

3.1 Data Mining in online shopping:

Data mining techniques are used in online shopping. When a customer is searching for some products to buy online then data mining technique can help in finding any possible relationship between different items. The discovery of such associations can help in increasing sales, resulting in increased profit.

3.2 Financial applications:

Organizations working in finance field such as banks, mutual fund companies, insurance companies etc. have to deal with huge amount of data every day. Data such as customer transactions, credit, loan, checks, savings accounts, stocks, various funds etc. is very crucial and important. Data mining techniques and data analytics proves helpful to deal with such huge data. Data mining techniques are helpful for increasing their business and also in fraud detection in some cases.

3.3 Telecom industry:

The telecommunication industry services are having various modes of communications to serve the customers such as telephone, mobile phone, fax, e-mail, internet etc. these different modes of communication generate a huge amount of data every day in the form of text, voice, images etc. This huge generation of data has boosted the use of

data mining techniques in telecom industry. The possible patterns can be identified using various data mining techniques resulting in better data management.

3.4 Text Mining:

As the name suggests text mining is the field where large amount of text documents are analyzed. Particular keywords and phrases are used to find any possible relationship or patterns in the documents.

3.5 Web Mining:

Web mining is used for websites. It is an emerging field which integrates data mining and text mining. When a customer is searching for a product online web mining techniques can find possible patterns which results in suggesting relevant links, new products. Previous customers with similar interests might have bought some other products. In such case, these products are suggested to the current customer.

3.6 Higher Education:

In today's era more and more students are aspiring higher education and specialization in particular field. To meet the needs of industry and research field there are variety of courses available. Data mining can be useful to the institutions and can better address the students and alumni through the analysis and presentation of data. Data mining has quickly emerged as a highly desirable tool for using current reporting capabilities to uncover and understand hidden patterns in vast databases.

3.7 Business applications:

The number of companies using data mining techniques to keep in pace with the competitive edge is increasing. These techniques are used in various business applications in investments, manufacturing, e-Commerce, targeted marketing etc.

3.8 Healthcare:

Data mining applications in healthcare can have huge scope. But it depends on the cleanliness of the healthcare data. To better utilize data mining techniques the data in health care industry should properly collected, store and maintain.

3.9 Bioinformatics:

Bioinformatics has a wide scope for using the data mining techniques. Some of the areas are gene finding, protein

function domain detection, disease diagnosis, disease treatment optimization, protein and gene interaction network reconstruction and protein sub-cellular location prediction.

CONCLUSION:

In this paper the some of the different tools, techniques and important applications of Data Mining are reviewed. It gives a researcher a brief idea about the multi-disciplinary tools, techniques and applications of this promising field at a glance.

REFERENCES:

1. Simmi Bagga, Dr. G.N. Singh, "Applications of Data Mining", International Journal of Science and Emerging Technologies with latest Trends, 1(1),2012, pp 19-23
2. Rayid Ghani, Carloas Saores, "Data Mining for business applications", SIGKDD Explorations, 2006, Volume 8, Issue 2, pp79-81
3. Annan Naidu Paidi, "Data Mining Future Trends and Applications", International Journal of Modern Engineering and Research, Volume, Issue 6, Nov-Dec 2012, pp 4657-4663, ISSN:2249-6645
4. Jiban K. Pal, "Usefulness and applications of data mining in extracting information from different perspectives", Annals of Library and information studies, Vol 58, march 2011, pp 7-16
5. Waldemar Wójcik and Konrad Gromaszek (2011). Data Mining Industrial Applications, Knowledge-Oriented Applications in Data Mining, Prof. Kimito Funatsu (Ed.), ISBN: 978-953-307-154-1, InTech, Available from:<http://www.intechopen.com/books/knowledge-oriented-applications-in-data-mining/data-mining-industrialapplications>
6. Cheng Yu, Xiong Ying, "Application of Data Mining Technology in E-Commerce", 2009 International Forum on Computer Science-Technology and Applications, 978-0-7695-3930-0/09 IEEE DOI 10.1109/IFCSTA.2009.76, pp 291-293
7. RupeshSanchati, P.C. Patidar, GauravKulkarni, "Path Breaking Case Studies in E-commerce using Data Mining ", International Journal of Computer Technology and Electronics Engineering (IJCTEE) Volume 1, Issue 1, pp 20-24
8. N R SrinivasaRaghavan, "Data mining in e-commerce: A survey", Sadhana Vol. 30, Parts 2 & 3, April/June 2005, pp. 275-289.
9. Fangfang Zhang, "The Application of Visualization Technology on E-commerce Data Mining", IEEE DOI 10.1109/IITA.2008.18
10. Benny Pinkas, "Cryptographic techniques for privacy-preserving data mining", SIGKDD Explorations, Volume 4, Issue 2Copyright Hewlett-Packard Company 2003, pp 1-14
11. QU Ziming, "Application of Grey Relational Clustering and Data Mining in Data Flow of Ecommerce", 978-0-7695-3645-3/09 IEEE DOI 10.1109/CINC.2009.57
12. Sunita Sarawagi et al, "Data mining models as services on the internet", proceedings of SIGKDD Explorations, ACM SIGKDD, July 2000pp 24-28