DATA MINING TREND IN PAST, CURRENT AND FUTURE

Article

2 authors, including:

DATA MINING TREND IN PAST, CURRENT AND FUTURE

*Sangeeta Goele, **Nisha Chanana
*Research Scholar, University School of Management, Kurukshetra University, Kurukshetra
**Research Scholar, University School of Management, Kurukshetra University
sangeetagoele04@gmail.com, nishachanana@gmail.com

ABSTRACT

Data and Information or Knowledge has a significant role on human activities. Data mining is the knowledge discovery process by analyzing the large volumes of data from various perspectives and summarizing it into useful information. Due to the importance of extracting knowledge/information from the large data repositories, data mining has become an essential component in various fields of human life including business, education, medical, scientific etc. Hence, this paper discusses the various improvements in the field of data mining from past to the present and explores the future trends.

INTRODUCTION

The data collected from different applications require proper mechanism of extracting knowledge/information from large repositories for better decision making. Knowledge discovery in databases (KDD), often called data mining, aims at the discovery of useful information from large collections of data. Data mining (sometimes called data or knowledge discovery) is the process of analyzing data from different perspectives and summarizing it into useful information - information that can be used to increase revenue, cuts costs, or both. Data mining software is one of a number of analytical tools for analyzing data. It allows users to analyze data from many different dimensions or angles, categorize it, and summarize the relationships identified. Technically, data mining is the process of finding correlations or patterns among dozens of fields in large relational databases.

This paper is organized as follows section 2 presents historical perspectives of data mining section 3 presents current trends in data mining section 4 presents future trends of data mining Section 5 presents the comparative statement of data mining trends and finally conclusion follows.

1. DEFINITION

"Data Mining represents a process developed to examine large amounts of data routinely collected. The term also refers to a collection of tools used to perform the process. Data mining is used in most areas where data are collected-marketing, health, communications, etc."

Example, one Midwest grocery chain used the data mining capacity of ORACLE software to analyze local buying patterns. They discovered that when men bought diapers on Thursdays and Saturdays, they also tended to buy beer. Further analysis showed that these shoppers typically did their weekly grocery shopping on

Saturdays. On Thursdays, however, they only bought a few items. The retailer concluded that they purchased the beer to have it available for the upcoming weekend. The grocery chain could use this newly discovered information in various ways to increase revenue. For example, they could move the beer display closer to the diaper display. And, they could make sure beer and diapers were sold at full price on Thursdays.

2. HISTORICAL TRENDS OF DATA MINING

Data mining is useful in various disciplines, which includes database management systems (DBMS), Statistics, Artificial Intelligence (AI), and Machine Learning (ML). The era of data mining applications was conceived in the year1980 primarily by research-driven tools focused on single tasks. The early day's data mining trends are as under.

2.1 Data Trends

In initial days, data mining algorithms work best for numerical data collected from a single data base, and various data mining techniques have evolved for flat files, traditional and relational databases where the data is stored in tabular representation. Later on, with the confluence of Statistics and Machine Learning techniques, various algorithms evolved to mine the non numerical data and relational databases.

2.2 Computing Trends

The field of data mining has been greatly influenced by the development of fourth generation programming languages and various related computing techniques. In, early days of data mining most of the algorithms employed only statistical techniques. Later on they evolved with various computing techniques like AI, ML and Pattern Reorganization. Various data mining techniques (Induction, Compression and Approximation) and algorithms developed to mine the large volumes of heterogeneous data stored in the data warehouses.

3. CURRENT TRENDS

The field of data mining has been growing due to its enormous success in terms of broad-ranging application achievements and scientific progress, understanding. Various data mining applications have been successfully implemented in various domains like health care, finance, retail, telecommunication, fraud detection and risk analysis etc. The ever increasing complexities in various fields and improvements in technology have posed new challenges to data mining; the various challenges include different data formats, data from disparate locations, advances in computation and networking resources, research and scientific fields, ever growing business challenges etc. Advancements in data mining with various integrations and implications of methods and techniques have shaped the present data mining applications to handle the various challenges, the current trends of data mining applications are:

3.1 Mining the Heterogeneous data

The following table depicts various currently employed data mining techniques and algorithms to mine the various data formats in different application areas. The various data mining areas are explained after the table1.

Table 1: Current Data Mining areas and techniques to mine the various Data format

Data mining type	Application Areas	Data Formats	Data mining Techniques/Algorithms
Hypermedia data mining	Internet and Intranet Applications.	Hyper Text Data	Classification and Clustering Techniques
Ubiquitous data mining	Applications of Mobile phones, PDA, Digital Cam etc.	Ubiquitous Data Traditional data mining techniques drawn from the Statistics and Machine Learning	Traditional data mining techniques drawn from the Statistics and Machine Learning
Multimedia data mining	Audio/Video Applications	Multimedia Data	Rule based decision tree classification algorithms
Spatial Data mining	Network, Remote Sensing and GIS applications.	Spatial Data	Spatial Clustering Techniques, Spatial OLAP
Time series Data mining	Business and Financial applications.	Time series Data	Rule Induction algorithms

3.2 Research and Scientific Computing Trends

The explosion in the amount data from many scientific disciplines, such as astronomy, remote sensing, bioinformatics, combinatorial chemistry, medical imaging, and experimental physics are tuning to various data mining techniques, to find out useful information. The Direct-kernel based techniques are powerful data mining tool for predictive modeling, feature selection and visualization in scientific computing.

3.3 Business Trends

Today's business must be more profitable, react quicker and offer high quality services that ever before. With these types of expectations and constraints, data mining becomes a fundamental technology in enabling customer's transactions more accurately. Data mining techniques of classification, regression, and cluster analysis are used for in current business trends. Most of the current business data mining applications utilize the classification and prediction techniques for supporting business decisions. In business environment data mining has evolved to Decision Support Systems (DSS) and very recently it has grown to Business Intelligence (BI) systems.

4. FUTURE TRENDS

Due to the enormous success of various application areas of data mining, the field of data mining has been establishing itself as the major discipline of computer science and has shown interest potential for the future developments. Ever increasing technology and future application areas are always poses new challenges and opportunities for data mining, the typical future trends of data mining includes:

- Standardization of data mining languages
- Data preprocessing
- Complex objects of data
- Computing resources
- Web mining
- Scientific Computing
- Business data

4.1 Standardization of data mining languages:

There are various data mining tools with different syntaxes, hence it is to be standardized for making convenient of the users. Data mining applications has to concentrate more in standardization of interaction languages and flexible user interactions.

4.2 Data Preprocessing:

To identify useful novel patterns in distributed, large, complex and temporal data, data mining techniques has to evolve in various stages. The present techniques and algorithms of data preprocessing stage are not up to the mark compared with its significance in finding out the novel patterns of data. In future there is a great need of data mining applications with efficient data preprocessing techniques.

4.3 Complex object of data:

Data mining is going to penetrate in all fields of human life; the presently available data mining techniques are restricted to mine the traditional forms of data only, and in future there is a potentiality for data mining techniques for complex data objects like high dimensional, high speed data streams, sequence, noise in the time series, graph, Multi-instance objects, Multi-represented objects and temporal data.

4.4 Computing Resources:

The contemporary developments in high speed connectivity, parallel, distributed, grid and cloud computing has posed new challenges for data mining. The high speed internet connectivity has posed a great demand for novel and efficient data mining techniques to analyze the massive data which is captured of IP packets at high link speeds in order to detect the Denial of Service (DoS) and other types of attacks.

Distributed data mining applications demand new alternatives in different fields, such as discovery of universal strategy to configure a distributed data mining, data placement at different locations, scheduling, resource management, and transactional systems etc. New data mining techniques and tools are needed to facilitate seamless integration of various resources in grid based environment. Moreover, grid based data mining has to focus seriously to address the data privacy, security and governance. Cloud computing is a great area to be focused by data mining, as the Cloud computing is penetrating more and more in all ranges of business and scientific computing. Data mining techniques and applications are very much needed in cloud computing paradigm.

4.5 Web mining:

The development of World Wide Web and its usage grows, it will continue to generate ever more content, structure, and usage data and the value of Web mining will keep increasing. Research needs to be done in developing the right set of Web metrics, and their measurement procedures, extracting process models from usage data, understanding how different parts of the process model impact various Web metrics of interest, how the process models change in response to various changes that are made-changing stimuli to the user, developing Web mining techniques to improve various other aspects of Web services, techniques to recognize known frauds and intrusion detection.

4.6 Scientific Computing:

In recent years data mining has attracted the research in various scientific computing applications, due to its efficient analysis of data, discovering meaningful new correlations, patterns and trends with the help of various tools and techniques. More research has to be done in mining of scientific data in particular approaches for mining astronomical, biological, chemical, and fluid dynamical data analysis. The ubiquitous use of embedded systems in sensing and actuation environments plays major impending developments in scientific computing will require a new class of techniques capable of dynamic data analysis in faulty, distributed framework. The research in data mining requires more attention in ecological and environmental information analysis to utilize our natural environment and resources. Significant data mining research has to be done in molecular biology problems.

4.7 Business Trends:

Business data mining needs more enhancements in the design of data mining techniques to gain significant advantages in today's competitive global market place (E-Business). The Data mining techniques hold great promises for developing new sets of tools that can be used to provide more privacy for a common man,

increasing customer satisfaction, providing best, safe and useful products at reasonable and economical prices, in today's E-Business environment.

5. COMPARATIVE STATEMENT

Table 2 present the comparative statement of various data mining trends from past to the future. Table describes the techniques, formats and resources used in different applications in past, current and future and shows with the change in time data mining techniques are improved and used in every area of industries.

Table 2: Data mining Trends comparative statements

Data mining trends	Algorithms/ Techniques employed	Data formats	Computing Resources	Prime areas of applications
Past	Statistical, Machine Learning Techniques	Numerical data and structured data stored in traditional databases	Evolution of 4G PL and various related techniques	Business
Current	Statistical, Machine Learning, Artificial Intelligence, Pattern Reorganization Techniques	Heterogeneous data formats includes structured, semi structured and unstructured data	High speed networks, High end storage devices and Parallel, Distributed computing etc…	Business, Web, Medical diagnosis etc.
Future	Soft Computing techniques like Fuzzy logic, Neural Networks and Genetic Programming	Complex data objects includes high dimensional, high speed data streams, sequence, noise in the time series, graph, Multi instance objects, Multi represented objects and temporal data etc…	Multi-agent technologies and Cloud Computing	Business, Web, Medical diagnosis, Scientific and Research analysis fields (bio, remote sensing etc.), Social networking etc

6. CONCLUSION

In this paper we try to briefly review the various data mining trends from its inception to the future. This review would be helpful to researchers to focus on the various issues of data mining. We found that Data mining is becoming increasingly common in both the private and public sectors. Industries such as banking, insurance, medicine, and retailing commonly use data mining to reduce costs, enhance research, and increase sales. So, data mining will be more and more useful in future.

7. REFERENCES

Books & Journals:

Distributed Processing Symposium (IPDPS'05). Gareth Herschel (1 July 2008) Magic Quadrant for Customer Data-Mining Applications

Ian H. Witten; Eibe Frank; Mark A. Hall (30 January 2011). Data Mining: Practical Machine Learning Tools and Techniques (3 ed.).

Jing He.2009. Advances in Data Mining: History and Future, Third international Symposium on Information Technology.

Karl Rexer, Heather Allen, & Paul Gearan (2010) 2010 Data Miner Survey Summary, presented at Predictive Analytics World, Oct. 2010.

Kotsiantis, S., Kanellopoulos, D., Pintelas, P. 2004. Multimedia Mining. WSEAS Transactions on Systems, No 3, s. 3263-3268.

Ma, Y.; Guo, Y.; Tian, X.; Ghanem, M. (2011). "Distributed Clustering-Based Aggregation Algorithm for Spatial Correlated Sensor Networks". IEEE Sensors Journal **11** (3): 641.

Piatetsky-Shapiro, Gregory. 2000. The Data-Mining Industry Coming of Age. IEEE Intelligent Systems.

Shonali Krishnaswamy. 2005. Towards Situationawareness and Ubiquitous Data Mining for Road Safety: Rationale and Architecture for a Compelling Application (2005), Proceedings of Conference on Intelligent Vehicles and Road Infrastructure 2005, pages-16, 17.

S. Mitra, S. K. Pal, and P. Mitra. 2001. Data mining in soft computing framework: A survey, IEEE Trans. Neural Networks, vol. 13, pp. 3 - 14.

Z. K. Baker and V. K.Prasanna. 2005. Efficient Parallel Data Mining with the Apriori Algorithm on FPGAs.

Websites:

http://www.alchemyapi.com

www.cs.umn.edu/~kumar/dmbook

http://www.kdnuggets.com/

www.loginworks.com/web-data-mining

www.mozenda.com/

www.smartertools.com/smarterstats/website-data-mining.aspx

www.web-datamining.net