IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 5, Issue 9, September 2016

Data Mining: Current Trends and Applications

Sandeep Godara[1]

Abstract: Data mining have various applications and these applications have enriched the various fields of human life including business, education, medical, scientific etc. There is formerly a blast in the amount of data which is produced and have accessed. Data mining can unearth significant information and widespread utilizations in business, medical, telecommunication wireless and ubiquitous computing. Objective of this paper is to discuss various improvements and breakthroughs in the field of data mining also to explore the future trends.

Keywords: Data mining, Knowledge discovery, business, medical, telecommunication, wireless and ubiquitous computing

I. INTRODUCTION

Data mining is then the process of extracting out valid and yet previously unknown information from large databases and using it to make critical decisions [3].Data mining or exploratory data analysis with large and complex datasets brings together the wealth of knowledge and research in statistics and machine learning for the task of discovering new snippets of knowledge in very large databases. From the user's point of view, the four steps listed in Table 1 were revolutionary because they allowed new business questions to be answered accurately and quickly. The core components of data mining technology have been under development for decades, in research areas such as statistics, artificial intelligence, and machine learning. Today, the maturity of these techniques, coupled with high-performance relational database engines and broad data integration efforts, make these technologies practical for current data warehouse environments

Table: 1

Evolutionary Step	Business Question	Enabling Technologies	Product Providers	Characteristics
Data Collection (1960s)	"What was my total revenue in the last five years?"	Computers, tapes, disks	IBM, CDC	Retrospective, static data delivery
Data Access (1980s)	"What were unit sales in New England last March?"	Relational databases (RDBMS), Structured Query Language (SQL), ODBC	Oracle, Sybase, Informix, IBM, Microsoft	Retrospective, dynamic data delivery at record level
Data Warehousing &Decision Support (1990s)	"What were unit sales in New England last March? Drill down to Boston."	On-line analytic processing (OLAP), multidimensional databases, data warehouses	Pilot, Comshare, Arbor, Cognos, Microstrategy	Retrospective, dynamic data delivery at multiple levels

An Introduction to Data Mining. Pilot Software Whitepaper. Pilot Software. 1998 Knowledge discovery process transforms data into knowledge [Cios K.J., 2000]. The key issue in KDD process shown in Fig 1. is to realise that there is more information hidden in the data than what could be made out at first sight. Before one attempts to extract useful knowledge from data, it is important to understand the overall approach. The process of knowledge discovery in databases is defined as the nontrivial process of identifying valid, novel, potentially useful, and ultimately understandable patterns in data. The process generalises to non-database sources of data, although it emphasises databases as a primary source of data. It consists of many steps (one of them is DM), each attempting to complete a particular discovery task and each accomplished by the application of a discovery method. Knowledge discovery concerns the entire knowledge extraction process, including how data are stored and accessed, how to use efficient and scalable algorithms to analyse massive datasets, how to interpret and visualise the results, and how to model and support the interaction between humans and machines. It also concerns support for learning and analysing the application domain.

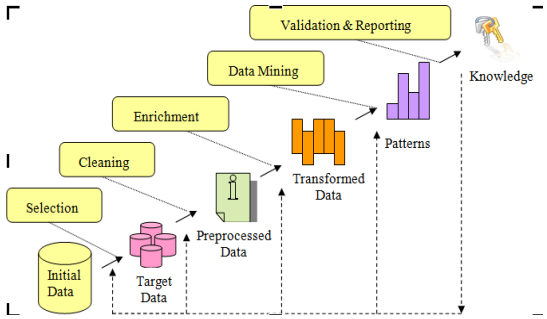

Figure 1: The process of knowledge discovery

This paper will discuss current trends and applications in Data Mining.

II. CURRENT TRENDS AND APPLICATIONS

In this section we will discuss applications of data mining.

A. Data mining in business

Business is a diverted field with several general areas of specialization such as accounting or financial analysis. Almost any neural network application would fit into one business area or financial analysis. There is some potential for using neural networks for business purposes, including

Copyright to IJARCCE

IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 5, Issue 9, September 2016

resource allocation and scheduling[1]. There is also a strong potential for using neural networks for database mining that is, searching for patterns implicit within the explicitly stored information in databases. There is a marketing application which has been integrated with a neural network system. The Airline Marketing Tactician is a computer system made of various intelligent technologies including expert systems[2]. A feed forward neural network is integrated with the AMT and was trained using back-propagation to assist the marketing control of airline seat allocations. The adaptive neural approach was amenable to rule expression. Additionally, the application's environment changed rapidly and constantly, which required a continuously adaptive solution. The system is used to monitor and recommend booking advice for each departure. Such information has a direct impact on the profitability of an airline and can provide a technological advantage for users of the system. [Hutchison & Stephens, 1987]

The HNC Company, founded by Robert Hecht-Nielsen, has developed several neural network applications. One of them is the Credit Scoring system which increases the profitability of the existing model up to 27%. The HNC neural systems were also applied to mortgage screening. A neural network automated mortgage insurance underwriting system was developed by the Nestor Company. This system was trained with 5048 applications of which 2597 were certified. The data related to property and borrower qualifications. In a conservative mode the system agreed on the underwritters on 97% of the cases. In the liberal model the system agreed 84% of the cases. This is system run on an Apollo DN3000 and used 250K memory while processing a case file in approximately 1 sec[2].

B. Telecommunication Industry

The telecommunication industry provides a lot of services in addition to telephone service i.e. fax, pager, cellular phone, Internet messenger, images, e-mail, computer and web data transmission etc. This has created a great demand for data mining in order to help understand the business involved, identify telecommunication patterns, catch fraudulent activities, make better use of resources, and improve the quality of service. The telecommunications industry generates and stores a tremendous amount of data. These data include call detail data, which describes the calls that traverse the telecommunication networks; network data, which describes the state of the hardware and software components in the network; and customer data, which describe the telecommunication customers.

The amount of data is so great that manual analysis of the data is difficult, if not impossible. Telecommunication data poses several interesting issues for data mining. The first concerns scale, since telecommunication databases may contain billions of records and are amongst the largest in the world. A second issue is that the raw data is often not suitable for data mining [3].

C. Fight against Terrorism

After 9-11 attacks, many countries imposed new laws against fighting terrorism. These laws allow intelligence agencies to effectively fight against terrorist organizations. USA launched Total Information Awareness program with the goal of creating a huge database of that consolidate all the information on population. Similar projects were also launched in European countries and rest of the world. This program faced several problems, a. The heterogeneity of database, the target database had to deal with text, audio, image and multimedia data. b. Second problem was scalability of algorithms. The execution time increases as size of data (which is huge). For example, 230 cameras were placed in London, to read number plates of vehicles. An estimated 40,000 vehicles pass camera every hour, in this way the camera must recognize 10 vehicles per second, which poses heavy load.

D. Biological Data Analysis

Recently, the collection of biological data has seen exponential increase due to improvements in existing technologies and the introduction of new ones such as the microarrays. These technological advances have assisted the conduct of large scale experiments and research programs. The explosive growth in the amount of biological data demands the use of computers for the organisation; the maintenance and the analysis of these recent data mining development tools assume importance in the analysis of data in order to gain new biological insights. The field of bioinformatics has many applications in the modern day world, including molecular medicine, industry, agriculture, stock farming, and comparative studies intense order calculations

E. Data Mining in Health Care Management

Nowhere in the field of science is the need for tools to deal with uncertainty more critical than in medicine, as disease diagnosis involves several levels of imprecision and uncertainty. A single disease may manifest itself quite differently in different patients and with different disease status. Further, a single symptom may be indicative of different diseases, and the presence of several diseases in a single patient may disrupt the expected symptom pattern of any of them [Massad E., 1999]. Clinical trails provide a method for evaluating the effectiveness and safety of new treatment for different diseases on human subjects [Bethel C.L. et.al. 2006]. Valuable knowledge can be discovered from the healthcare datasets by using the data mining techniques. Data mining and interactive decision support tools can help the clinicians to process a huge amount of data available from previous cases and help in diagnostic study to suggest probable ailment based on the values of several important attributes of patient

F. Spatial Data Mining

Spatial data mining is the application of data mining techniques to spatial data. Spatial data mining follows along the same functions in data mining, with the end objective to find patterns in geography. So far, data mining

ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 5, Issue 9, September 2016

and Geographic Information Systems (GIS) have existed as two separate technologies, each with its own methods, traditions and approaches to visualisation and data analysis. Particularly, most contemporary GIS have only very basic spatial analysis functionality. The immense explosion in geographically referenced data, digital mapping, remote sensing, and the global diffusion of GIS emphasises the importance of developing data driven inductive approaches to geographical analysis and modeling.The discovery process for spatial data is more complex than for relational data. The reason is that, in contrast to mining in relational databases, spatial data mining algorithms have to consider the neighbours of objects in order to extract useful knowledge. This applies to both the efficiency of algorithms as well to the complexity of possible patterns that can be found in a spatial database. In some spatial databases the dimension of time plays an important role, or example, raster images of the same area of the surface of the earth taken at different times. Data mining in such spatio-temporal databases is a promising area of research.

G. Ubiquitous Data Mining

Accessing and analyzing data from a ubiquitous computing device offer many challenges. For example, UDM introduces additional cost due to communication, computation, security, and other factors. So one of the Human-computer interaction is another challenging aspect of UDM. Visualizing patterns like classifiers, clusters, associations and others, in portable devices are usually difficult. The small display areas offer serious challenges to interactive data mining environments. Data management in a mobile environment is also a challenging issue. Moreover, the sociological and psychological aspects of the integration between data mining technology and our lifestyle are yet to be explored. The key issues to consider include theories of UDM, advanced algorithms for mobile and distributed applications, data management issues, mark-up languages, and other data representation techniques; integration with database applications for mobile environments.

H. Time Series Data Mining

Another important area in data mining centers on the mining of time series and sequence-based data. Simply put, this involves the mining of a sequence of data, which can either be referenced by time (time-series, such as stock market and production process data), or is simply a sequence of data which is ordered in a sequence. In general, one aspect of mining time series data focuses on the goal of identifying movements or components which exist within the data (trend analysis).

III. CONCLUSION

In closing, it would not be overly optimistic to say that data mining has a bright and promising future, and that the years to come will bring many new developments, methods, and technologies. In addition, the improved integration of techniques and the application of data mining techniques can bring about the handling of new kinds of data types and applications. By expanding applications which can use it, integrating technologies and methods, broadening its applicability to mainstream business applications, and making programs and interfaces easier for end-users to use, it is quite possible and likely that data mining will become one of the key technology areas of the new millennium.

REFERENCES

1. Huysmans, Baesens, Martens, Denys and Vanthienen, “New Trends in Data Mining”, Tijdschrift voor Economie en Management, vol. L, 4, 2005.
2. Kargupta, Han, Yu, Motwani, Vipin Kumar, “Next Generation of Data Mining”, Chapman & Hall /CRC Data Mining and Knowledge Discovery Series, Taylor and Francis Group LLC, 2008.
3. Hsu J., “Data Mining Trends and Developments: The Key Data Mining Technologies and Applications for the 21st Century”, in The Proceedings of the 19th Annual Conference for Information Systems Educators (ISECON 2002), ISSN: 1542-7382. Available Online: http://colton.byuh.edu/isecon/2002/224b/Hsu.pdf
4. Shonali Krishnaswamy, “Towards Situationawareness and Ubiquitous Data Mining for Road Safety: Rationale and Architecture for a Compelling Application”, Proceedings of Conference on Intelligent Vehicles and Road Infrastructure 2005, pages16,17.Availableathttp://www.csse.monash.edu.au/~mgaber/CameraReadyI
5. Sunila , Prabhat, and Nirmal Godara. "Decision support system for cardiovascular heart disease diagnosis using improved multilayer perceptron." *International Journal of Computer Applications* 45.8 (2012).
6. Sunila Godara, Amita Verma, "Analysis of Various Clustering Algorithms",International Journal of Innovative Technology and Exploring Engineering (IJITEE), ISSN: 2278-3075, Volume-3, Issue-1, Page 186-189 and June 2013,.
7. Sunila Godara, ., and R. Yadav. "Performance analysis of clustering algorithms for character recognition using Weka tool." *International Journal of Advanced Computer and Mathematical Sciences* 4.1 (2013): 119-123.
8. Sunila G, Rishipal S. "Machine Learning For Medical Decision Support Systems (MDSS): A review .",International Journal of Applied Engineering Research. 2015; 10(13):32864–73.
9. Sunila Godara , Rishipal Singh, "Evaluation of Predictive Machine Learning Techniques as Expert Systems in Medical Diagnosis", Indian Journal of Science and Technology, vol. 910, 2016.

DOI 10.17148/IJARCCE.2016.5926