Journal of Basic and Applied Engineering Research
Print ISSN: 2350-0077; Online ISSN: 2350-0255; Volume 2, Number 10; April-June, 2015 pp. 860-864

http://www.krishisanskriti.org/jbaer.html

Social Implications of Data Mining Techniques

Sangeeta Pathak

IME College, Sahibabad
E-mail: ssangeeta.ppathak@gmail.com

Abstract—*Data mining is the process of knowledge finding where knowledge is gained by analysing the data store in very large sources, which are analysed from various viewpoints and the result is summarized it into useful information. Due to the importance of extracting knowledge/information from the large data repositories, data mining has become a very important and guaranteed branch of engineering affecting human life in various spheres directly or indirectly. Advancements in Statistics, Machine Learning, Artificial Intelligence, Pattern recognition and Computation capabilities have given present day's data mining functionality a new height. Data mining have various applications and these applications have enriched the various fields of human life including business, education, medical, scientific etc. Objective of this paper is to discuss various improvements and breakthroughs in the field of data mining from past to the present and also to explores the future trends.*

1. INTRODUCTION

Data mining is the process of knowledge discovery where knowledge is gained by analysing the data store in very large repositories, which are analysed from various perspectives and the result is summarized it into useful information. Due to the importance of extracting knowledge/information from the large data repositories, data mining has become a very important and guaranteed branch of engineering affecting human life in various spheres directly or indirectly. Advancements in Statistics, Machine Learning, Artificial Intelligence, Pattern recognition and Computation capabilities have given present day's data mining functionality a new height. Data mining have various applications and these applications have enriched the various fields of human life including business, education, medical, scientific etc. Objective of this paper is to discuss various improvements and breakthroughs in the field of data mining from past to the present and also to explores the future.

For most of us, data mining has become a part of our daily lives. Data mining, affecting everyday things from the products stocked at our local supermarket, to the ads we see while surfing the Internet, to crime prevention. Data mining can offer the individual many benefits by improving customer service and satisfaction, and lifestyle, in general.Data mining is present in many aspects of our daily lives, whether we realize it or not. Itaffects how we shop, work, search for information, and can even influence our leisure time, health, and well-being. In this section, we look at examples of such ubiquitous (or ever-present) data mining. Several of these examples also represent invisible data mining, in which "smart" software, such as Web search engines, customer-adaptive Web services (e.g., using recommender algorithms), "intelligent" database systems, e-mail managers, ticket masters, and so on, incorporates data mining into its functional components, often unbeknownst to the user.

2. CURRENT TRENDS OF DATA MINING

2.1 E-Shopping

Data mining has innovatively influenced what we buy, the way we shop, as well as our experience while shopping. One example is Wal-Mart, which has approximately 100 million customers visiting its more than 3,600 stores in the United States every week. Wal-Mart has 460 terabytes of point-of-sale data stored on Teradata mainframes, made by NCR. To put this into perspective, experts estimate that the Internet has less than half this amount of data. Wal-Mart allows suppliers to access data on their products and perform analyses using data mining software. This allows suppliers to identify customer buying patterns, control inventory and product placement, and identify new merchandizing opportunities. All of these affect which items (and how many) end up on the stores' shelves—something to think about the next time you wander through the aisles at Wal-Mart.

Data mining has shaped the on-line shopping experience. Many shoppers routinely turn to on-line stores to purchase books, music, movies, and toys. The use of collaborative recommender systems, which offer personalized productrecommendations based on the opinions of other customers. Amazon.com was at theforefront of using such a personalized, data mining–based approach as a marketingstrategy. CEO and founder Jeff Bezos had observed that in traditional brick-and-mortar stores, the hardest part is getting the customer into the store. Once the customer is there, she is likely to buy something, since the cost of going to another store is high. Therefore, the marketing for brick-and-mortar stores tends to emphasize drawing customers in, rather than the actual in-store customer experience. This is in contrast to on-line stores, where customers can "walk out" and enter another on-line store with just a click of the mouse. Amazon.com capitalized on this difference, offering a

"personalized store for every customer." They use several data mining techniques to identify customer's likes and make reliable recommendations.

2.2 Customer relationship management(CRM)

Many companies increasingly use data mining for customer relationship management(CRM), which helps provide more customized, personal service addressing individual customer's needs, in lieu of mass marketing. By studying browsing and purchasing patterns on Web stores, companies can tailor advertisements and promotions to customer profiles, so that customers are less likely to be annoyed with unwanted mass mailings or junk mail. These actions can result in substantial cost savings for companies. The customers further benefit in that they are more likely to be notified of offers that are actually of interest, resulting in less waste of personal time and greater satisfaction. This recurring theme can make its way several times into our day, as we shall see later.

2.3 Search Engines

Data mining has greatly influenced the ways in which people use computers, search for information, and work. Suppose that you are sitting at your computer and have just logged onto the Internet. Chances are, you have a personalized portal, that is, the initial Web page displayed by your Internet service provider is designed to have a look and feel that reflects your personal interests. Yahoo (*www.yahoo.com*) was the first to introduce this concept. Usage logs from My Yahoo are mined to provide Yahoo with valuable information regarding an individual's Web usage habits, enabling Yahoo to provide personalized content. This, in turn, has contributed to Yahoo's consistent ranking as one of the top Web search providers for years, according to *Advertising Age's BtoB*magazine's Media Power 50 (*www.btobonline.com*), which recognizes the 50 most powerful and targeted business-to-business advertising outlets each year.

After logging onto the Internet, you decide to check your e-mail. Unbeknownstto you, several annoying e-mails have already been deleted, thanks to a spam filter that uses classification algorithms to recognize spam. After processing your e-mail,you go to Google (*www.google.com*), which provides access to information from over2 billion Web pages indexed on its server. Google is one of the most popular and widely used Internet search engines. Using Google to search for information has become a way of life for many people. Google is so popular that it has even become a new verb in the English language, meaning "to search for (something) on the Internet using the Google search engine or, by extension, any comprehensive search engine."You decide to type in some keywords for a topic of interest. Google returns a list of websites on your topic of interest, mined and organized by PageRank. Unlike earlier search engines, which concentrated solely onWeb content when returning the pages relevant to a query, PageRank measures the importance of a page using structural link information fromthe Web graph. It is the core of Google's Web mining technology.

While you are viewing the results of your Google query, various ads pop up relating to your query. Google's strategy of tailoring advertising to match the user's interests is Successful—it has increased the clicks for the companies involved by four to five times. This also makes you happier, because you are less likely to be pestered with irrelevant ads. Google was named a top-10 advertising venue by Media Power 50.

Web-wide tracking is a technology that tracks a user across each site she visits. So, while SurfingtheWeb, informationabout every site youvisitmaybe recorded,whichcanprovide marketers with information reflecting your interests, lifestyle, and habits. DoubleClick Inc.'s DART ad management technology uses Web-wide tracking to target advertising based on behavioural or demographic attributes. Companies pay to useDoubleClick's service on their websites. The clickstream data from all of the sites using DoubleClick are pooled and analysed for profile information regarding users who visit any of these sites. DoubleClick can then tailor advertisements to end users on behalf of its clients. In general, customer-tailored advertisements are not limited to ads placed onWeb stores or company mail-outs. In the future, digital television and on-line books and newspapers may also provide advertisements that are designed and selected specifically for the given viewer or viewer group based on customer profiling information and demographics.

While you're using the computer, you remember to go to eBay (*www.ebay.com*) to see how the bidding is coming along for some items you had posted earlier this week. You are pleased with the bids made so far, implicitly assuming that they are authentic. Luckily, eBay now uses data mining to distinguish fraudulent bids from real ones.

2.4 Integration of data mining into existing business technology

Data mining and OLAP technologies can help us in our work in many ways. Business analysts, scientists, and governments can all use data mining to analyse and gain insight into their data. They may use data mining and OLAP tools, without needing to know the details of any of the underlying algorithms. All that matters to the user is the end result returned by such systems, which they can then process or use for their decision making. Data mining technologies are being used in business in many ways like, User Security, Inventory and Order Management System and Product Management etc.

2.5 Leisure time

Data mining can also influence our leisure time involving dining and entertainment. Suppose that, on the way home from work, you stop for some fast food. A major fast-food restaurant used data mining to understand customer behaviour

via market-basket and time-series analyses. Consequently, a campaign was launched to convert “drinkers”to “eaters” by offering hamburger-drink combinations for littlemore than the price of the drink alone. That’s food for thought, the next time you order a meal combo. With a little help fromdata mining, it is possible that the restaurant may even know what you want toorder before you reach the counter. Bob, an automated fast-food restaurant management system developed by HyperActive Technologies (*www.hyperactivetechnologies.com*), predictswhatpeople are likely toorder basedonthe typeof car they drive tothe restaurant, and on their height. For example, if a pick-up truck pulls up, the customer is likely to order quarter pounder. Afamily car is likely to include children, which means chicken nuggets and fries. The idea is to advise the chefs of the right food to cook for incoming customers to provide faster service, better-quality food, and reduce food wastage. After eating, you decide to spend the evening at home relaxing on the couch. Blockbuster(*www.blockbuster.com*) uses collaborative recommender systems to suggest movie rentals toindividual customers. Othermovie recommender systemsavailableonthe Internet includeMovie Lens (*www.movielens.umn.edu*)Netflix (*www.netflix.com*). (There are even recommender systems for restaurants,music, and books that are not specifically tied to any company.) Or perhaps you may prefer to watch television instead. NBC uses data mining to profile the audiences of each show. The information gleaned contributes toward NBC’s programming decisions and advertising. Therefore, the time and day of week of your favourite show may be determined by data mining.

2.6 Health and Well Being

Finally, data mining can contribute toward our health and well-being. Several pharmaceutical companies use data mining software to analyse data when developing drugs and to find associations between patients, drugs, and outcomes. It is also being used to detect beneficial side effects of drugs. The hair-loss pill Propecia, for example, was first developed to treat prostrate enlargement. Data mining performed on a study of patients found that it also promoted hair growth on the scalp. Data mining can also be used to keep our streets safe.

2.7 Crime Prevention

The data mining system Clementine from SPSS is being used by policedepartments to identify key patterns in crime data. It has also been used by police todetect unsolved crimes that may have been committed by the same criminal. Many policedepartments around the world are using data mining software for crime prevention, such as the Dutch police’s use of Data Detective (*www.sentient.nl*) to find patterns in criminal databases. Such discoveries can contribute toward controlling crime.

As we can see, data mining is omnipresent. For data mining to become further accepted and used as a technology, continuing research and development are needed in the many areas.

Major challenges are efficiency and scalability, increased user interaction, incorporation of background knowledge and visualization techniques, the evolution of a standardized data mining query language, effective methods for finding interesting patterns, improved handling of complex data types and stream data, real-time data mining, Web mining, and so on.

In addition, the *integration* of data mining into existing business and scientific technologies, to provide domain specific data mining systems, will further contribute toward the advancement of the technology.

3. 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.

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

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 pre-processing
- 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 Pre-processing

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 pre-processing 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 pre-processing 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.

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.

4.4 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.

4.5 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.

4.6 Sentiment Orientation (SO)

Widespread products are likely to attract thousands of reviews and this may make it difficult for prospective buyers to track usable reviews that may assist in making decision. On the other hand sellers make use of Sentiment Orientation (SO) for their rating standard in other to safeguard irrelevant or misleading reviews present to reviewers.. Live journal blog corpus dataset was used to train and evaluate the method used. The experiment presented a modular proficient hierarchical classificationtechnique easily implemented together with SO attributes and machine learning techniques. The initial result of classification accuracy however recorded slightly above the baseline. An incorporation of flexible hierarchy basedmood approach to mood classification discovers that attributes that points to accurate classification of mood expression can be retrieved from the various dense blog corpus domain.

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.

4.8 Comparative Statement

The following table 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

5. CONCLUSION

The success of data mining solutions tailored for e-commerce applications, as opposed to generic data mining systems, is an example. With more and more information accessible in electronic forms and available on the Web, and with increasingly powerful data mining tools being developed and put into use, there are increasing concerns that data mining may pose a threat to our privacy and data security. However, it is important to note that most of the major data mining applications do not even touch personal data. Prominent examples include applications involving natural resources, the prediction of floods and droughts, meteorology, astronomy, geography, geology, biology, and other scientific and engineering data.

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.

REFERENCES

[1] Distributed Processing Symposium (IPDPS'05). Gareth Herschel (1 July 2008) Magic Quadrant for Customer Data-Mining Applications
[2] Ian H. Witten; Eibe Frank; Mark A. Hall (30 January 2011). Data Mining: Practical Machine Learning Tools and Techniques (3 ed.).
[3] Jing He.2009. Advances in Data Mining: History and Future, Third international Symposium on Information Technology.

Websites

[1] www.loginworks.com/web-data-mining
[2] www.smartertools.com/smarterstats/website-data-mining.aspx
[3] www.web-datamining.net