

Data Mining, Knowledge Discovery and its Applications

Sanjuktaranijena,
Assistant Professor,
Department of Management Studies,
KMM Institute of Technology and Science,

S. Ismail Basha,
Assistant Professor,
K.M.M.Institute of Post Graduate Studies,
Department of Computer Science, Tirupati-517502,

Abstract-Data Mining and Knowledge Discovery is a trialed peer-reviewed scientific journal focusing on data mining. It is published by Springer Science + Business Media. As of 2012, the editor-in-chief is Geoffrey I. Webb.

The practice of data mining or knowledge discovery, despite highly intricate procedures and applications, is based on a very simple concept. Collecting information from a breath of sources for the purposes of analysis. Generally, data mining of hard drive recovery is done in order to glean information which can then be utilized to improve a process or technique. From business to science and data mining has become essential to standard operations. For an more in depth view of data mining and knowledge discovery and the variety of way in which it is used, visit the listing below for a full tour of this multi-faceted science.

Data mining and Knowledge Discovery has several important application areas. Data mining and Knowledge Discovery have been topics considered at many AI (Artificial Intelligence), database and statistical conferences. Knowledge discovery generally refers to the process of identifying valid, novel and understandable patterns. Knowledge discovery from large data bases or data sets. The discovery process can be broken into the several steps, including: developing an understanding of the application domain creating a target data set data clearing and processing finding useful features with which to represent the data; data mining to search for patterns of interest; and interesting and consolidating discovered patterns.

Keywords: Data Mining, Data Mining and KDD, How does data mining work? What technologies infrastructure is required?

I. INTRODUCTION

Generally, data mining (sometimes called data or knowledge discovery) is the process of analyzing data item from different perspectives and summarizing it into useful information that can be used to increase revenue, cuts costs, or both. Data mining software is one of a number of analytic tools for analyzing data. It allows users to analyze data from many different dimensions of 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 recreational data bases.

Data mining is a logical process that is used to search through large amounts of information order to find important data. The goal of this technique is to find

patterns that were previously unknown. Once you have found these patterns, you can use them to solve a number of problems.

Data mining is a powerful tool because it can provide you with relevant information that you can use to your own advantage. When you have the right knowledge, all you will need to do is apply it in the right manner, and you will be able to benefit. It is relatively easy to get information these days. This is where data mining becomes a powerful tool that you will want to become familiar with. It will give you the power to predict certain behaviors within a system.

Data mining has been defined in almost as many ways as there are authors who have written about it. Because it sits at the interface between statics, computer science, artificial intelligence, machine learning, database management and data visualization, the definition changes with the perspective of the user: "Data Mining is the process of exploration and analysis by automatic or semiautomatic means, of larger quantities of data in order to discover meaningful patterns and rules."(M.J.A.Berry and G.S. Linoff) Data Mining is finding interesting structure (patterns, statistical models, relationships) in databases".(U.Fayyad, S.Chaudhuri and P.Bradley).

II. DATA MINING AND KNOWLEDGE DISCOVERY IN DATABASE (KDD):

"Data Mining is the application of statistics in the form of exploratory data analysis and predictive models to reveal patterns and trends in very large data sets. The traditional method of turning data into knowledge relies on manual analysis and interpretation. For example, in the health-care industry, it is common for specialists to periodically analyze current trends and change sin health-care data, say, on a quarterly basis. The specialists then provide a report detailing the analysis to the sponsoring health-care management. In a totally different type of application, planetary geologists sift through remotely sensed images of planet sand asteroids, carefully locating and cataloging such geologic objects of interests as impact raters. Be it science, marketing, finance, health care, retail, or any other field, the classical approach to data analysis relies fundamentally on one or more analysis be coming intimately familiar with the data and servings an interface and between the data and the users and products.

Data Mining and Knowledge Discovery in the Real World. A large degree of the current interest in KDD is the result of the media interest surrounding successful **KDD applications**, for example, therefore articles within the last two years in Business Week, Byte, PC Week, and other large-circulation periodicals. Unfortunately, it is not always easy to separate fact from media hype. Nonetheless, several well documented examples of successful system scan rightly are referred to as **KDD application** sand have been deployed in operational use on large-scale real-world problems in science and in business. In science, one of the primary application areas is astronomy. In business, main **KDD application** areas includes marketing, finance (especially investment), fraud detection, manufacturing, telecommunications, and Internet agents.

Advance in data gathering storage and distribution have created a need for computational tools and techniques to aid in data analysis. Data Mining and Knowledge Discovery in Database (KDD) is a rapidly growing area of research and application that builds on techniques and theories from many fields including statistics databases pattern recognition and learning data visualization uncertainty modeling data warehousing and OLAP optimization and high performance computing KDD is concerned with issues of scalability the multi-step knowledge discovery process for extracting useful patterns and models from raw data stores (including data cleaning and noise modeling) and issues of making discovered patterns understandable. Data

III.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

IV.HOW DOES DATA MINING WORK?

Data mining software analyses relationships and patterns in this stored transaction data. Several types of analytical software are available: statistical, machine learning, and neural networks. Generally, any of four types of relationships are sought:

- **Classes:** Stored data is used to locate data in predetermined groups. For example, a restaurant chain could mine customer purchase data to determine when customers visit and what they

typically order. This information could be used to increase traffic by having daily specials. Data classes are groups that share easily identifiable characteristics. This explains why they are also referred to as **predetermined groups**.

In the context of a retail business, customers who have purchased a particular product constitute a data class.

For example, Amazon.com customers who have purchased business books in the past constitute a class. Knowing the characteristics of the data class takes the guesswork out of “likelihood to buy” factor in sales promotion. The online retailer can use this grouping to develop marketing campaigns for business books and target customers in the group (and underlying sub-groups). Depending on the size of each class, data grouping can significantly improve the efficiency of mass marketing.

- **Clusters:** Data items are grouped according to logical relationships or consumer preferences. For example, a sports shop that analyzed their data know that there is an 85% chance that a person buying new mountain bike will also buy a helmet, gloves and a water bottle. However, customers who come in requesting a helmet will probably not buy a bike, but they most likely will also buy gloves. This knowledge can assist the manager in ordering the correct stock and assist the sale personnel in suggesting add-on purchasing.

Data clusters are similar to classes, but include additional attributes such as logical relationships. In the context of business applications, consumer preferences are often the most useful attributes. Consumer preferences can be used to understand market segments and customer loyalty. Accurate clustering can support cross selling. Again, using Amazon.com as an example, data clusters allow the retailer to identify what other products are purchased by customers who buy business books. Armed with this information, the retailer can develop “product recommendations” as part of its customer relations management (CRM) programs. The ability to nurture leads efficiently is critical to sales.

- **Associations:** Data can be mined to identify associations. Data associations take clusters further. In the context of business application, associative data mining reveals buying patterns that would otherwise go unnoticed. For example, changes in buying habits induced by shifts in the economy require in-depth analysis for accurate characterization. A clear understanding of the economic shifts can be exploited for marketing purposes.

- **Sequential patterns:** Data is mined to anticipate behavior patterns and trends. For example, an outdoor equipment retailer could predict the likelihood of a backpack being purchased based on a consumer's purchase of sleeping bags and hiking shoes.

While analyzing past purchases is helpful, some experts believe that the true benefit of data mining is to anticipate customer purchases through predictive analytics. By building on historical data, sequential patterns allow projections to be developed. The projected industry trends are essential for forward-looking business planning and competitive intelligence.

V. DATA MINING CONSISTS OF FIVE MAJOR ELEMENTS:

- Extract, transform, and load transaction data onto the data warehouse system.
- Store and manage data in a multidimensional database system.
- Provide data access to business analysts and information technology professionals.
- Analyze data by application software.
- Present data in a useful format, such as a graph or table.

VI. CONCLUSION

Few will contest the potential of data mining tools to create valuable business insights. However, as with all technologies, the deployment of data mining needs to be driven by well-researched enterprise needs, as well as cost and usability considerations.

Without specific effort, your mind is building clusters and associations. When you see a man and a woman walking close to each other, you just know that they are either related or a couple. You see a woman coming out of a certain shop and you immediately associate her with the image the shop portrays.

Data mining systems just make it easier for us to handle large amounts of data. Almost everything that is done in data mining can be done manually by a human but that would just take tremendously longer.

REFERENCES

Books & Journals:

1. Distributed Processing Symposium (IPDPS'05). Gareth Herschel (1 July 2008) Magic Quadrant for Customer.
2. Data-Mining Applications.
3. Ian H. Witten; Eibe Frank; Mark A. Hall (30 January 2011). Data Mining: Practical Machine Learning Tools and Techniques (3 ed.).
4. Jing He. 2009. Advances in Data Mining: History and Future, Third international Symposium on Information Technology.
5. Karl Rexer, Heather Allen, & Paul Gearan (2010) 2010 Data Miner Survey Summary, presented at Predictive
6. Analytics World, Oct. 2010.
7. Kotsiantis, S., Kanellopoulos, D., Pintelas, P. 2004. Multimedia Mining. WSEAS Transactions on Systems, No 3, s. 3263-3268.
8. 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.
9. Piatetsky-Shapiro, Gregory. 2000. The Data-Mining Industry Coming of Age. IEEE Intelligent Systems.
10. Shonali Krishnaswamy. 2005. Towards Situationawareness and Ubiquitous Data Mining for Road Safety:
11. Rationale and Architecture for a Compelling Application (2005), Proceedings of Conference on Intelligent

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