

A Study on the Evolution of Data Mining Technologies

Mohd. Furkan

Research Scholar (Computer Science), CMJ University, Shillong, Meghalaya

Abstract — *Data mining happens to be a powerful technology which helps the companies to focus on important information which are lying in their data warehouses. They can answer questions in the business process which normally took a long time to solve earlier. In the world of today most of the companies collect large amount of information which can be implemented on the existing software as well as hardware platforms. Data mining provides answers to questions like “which clients are more likely to respond to the promotional campaigns”. The paper below provides into the insight of the various components of data mining technology, the importance of it along with the diverse areas where it finds its use.*

Index Terms— *Data Mining.*

1. INTRODUCTION

Data mining means searching information from a large database. The core components related to data mining have been used in diverse fields of statistics as well as artificial intelligence. Data mining technologies can generate new business opportunities. The progress in digital technology as well as storage technology has resulted in the growth of large databases. This has occurred in all the areas of human endure such as supermarkets, credit card usage astronomical as well as scientific data. So no wonder a lot of interest is shown in tapping this data.

Data mining typically involves collection of data which have already been collected other than the purpose of data analysis. For example they may have been collected to maintain the up to date record of transactions in a bank. This means that the object of data mining exercise play no role in the data collection strategy. The definition also means that data collected in data mining involves large sets of data. If only small sets of data were involved we would be dealing with the component of statistics. In fact the relationship between the various structures in a data must be novel. There is little point in segregating well established relationships within the data.

Data mining is the most powerful resource to meet the challenges of business. Computers are a part and parcel of our lives and are no longer considered a luxury item like the earlier days. The growth of information technology has

resulted in widespread use of computers in engineering as well as scientific streams. Data is the input which exists in the raw form which is used for further processing. With huge amount of data at their disposal organizations are facing a challenge to extract meaningful information from them. This leads to the emergence of the concept of data mining. In computer terminology data mining operates on raw input data and produces the output. In fact the entire crux of the data mining process could be referred to as the mechanism of transformation just like any other computer applications. On the other hand data mining could also be treated as an intelligent application system. This is the precise reason why it is viewed as one of the major classifications under the subset of artificial intelligence. In short they project the information which is available in a database (Gopala and Sivaselvan, 2009).

2. WHAT IS DATA MINING?

Data mining is one step in the data mining process. The definition of data mining as well as data mining process varies across both the cultures (Sumathi and Sivanandam, 2009).

Data mining is the automated discovery of associations, clusters and patterns in large data sets with the objective of attaining meaningful information from it. Data mining consists of a series of steps which are repeated in a sequential fashion. A glance at the steps are as follows

- Data preparation and cleaning

- Data warehousing
- Identifying predictive patterns
- Computing derived attributes
- Data reduction
- Modelling
- Knowledge extraction
- Interactive data analysis and recovery

The field of data mining is relatively new and lot of evolutions have taken place in this domain. In fact the first international conference on knowledge discovery as well as data mining (KDD) was held in the year 1995 and a variety of definitions have emerged ever since then. Data mining is used in a variety of field as well as applications. For example military may use data mining to determine the accuracy factor relating to the bombs whereas medical researches may use it to determine the likelihood of a cancer. Some common business questions which could be answered by the method of data mining are as follows

- From a prospective list of customers which ones are the most likely to respond to a particular marketing campaign. In this regard techniques like classification and decision trees may be used to determine which individuals demographic and other data matches with the best existing customers
- Which loan customers are likely to commit default in payments. In this regard we can use the classification technique to classify them
- Which customers are more likely to commit fraud or who have already committed fraud at some point of time.
- Which customers are more likely to abandon a subscription service. In this regard we can use a classification technique to determine them also.

Perhaps the most compelling reason for the growth of data mining is the growth of data (Shmueli, Patel and Bruce, 2010).

The growth of internet has created a new arena for information generation. Data from the operational databases are extracted, transformed and exported to a data warehouse. In fact many of the mechanisms used in data mining would not be possible without the growth of computational power in the world of today. So it is quiet clear that it is here to stay big time.

3. EVOLUTION OF DATA MINING TECHNOLOGIES

The techniques of data mining are designed and work with large data sets. The evolution of data mining first began when data was stored in computers and technologies were generated to allow the users to navigate through the data in real time. The evolution is possible because of the support of three major technologies, the massive data collection, high performance computing as well as data mining algorithms. It is a wider component of a wide process known as knowledge discovery from databases. In fact it involves personnel from various disciplines including mathematicians as well as scientists.

Data mining techniques are a result of a long process of research along with product development. The commercial databases are growing at an unprecedented rate in the world of today. Data mining techniques algorithms though have existed for the last ten years but have recently being implemented as a mature, reliable as well as an understandable tool that outweighs the traditional statistical methods. In the evolution of business data to business information each step is build upon the previous one. An example will illustrate this on a better level as data access is critical for drill through in data navigation applications and the ability to store data in large databases is a critical factor for data mining (Agarwal and Tayal, 2009).

Some of the commonly used tools in data mining are as follows

- Artificial neural networks- the nonlinear predictive models that incorporate the learning process through training and resemble biological neural networks in stature
- Decision trees- tree shaped trees that replicate decision making patterns. These decisions help in classification of a subset
- Data visualization- The visual presentation of complex relationships in data.
- Rule indication- The extraction of useful if then rules from data based on numerical significance (Sumathi and Sivanadam, 2011)

Data mining models are deployed for a purpose. The tools should allow the user to apply the model to the purpose. For example if the model is developed with the objective of identifying the best customers for a mailing campaign then the model should be available when the list is in build. Another approach to data mining is to develop data mining models through a service oriented interface. For example several models may be building to understand the customer segmentation and the models can be applied

through the server and any applications can incorporate and work on these models.

No doubt data mining has gone through an evolution since it was developed in the period of 1980's. Over the years data mining has been used to describe signatures, patterns as well as profiles in large databases through such process like segmentation, classification as well as prediction. Data mining is also descriptive in nature as it is about discovering signature in databases through the following patterns of the data mining process which are

- Classification with neural networks
- Clustering with self organizing maps
- Profiling with machine learning algorithms

Prior to the growth of the internet and the communications network data mining was at the infancy stage. In fact at the first stage of the evolution data has to be imported in the data mining software or linked to tables as well (Mena, 2007).

Given the level of the society's dependence on computer technology, it is quite reasonable to believe that the dependence on data mining in investigations is bound to rise all the more (Skalak, et al, 2011)

The database driven models as well as the subsequent prediction systems should be employed in environments where the models are constantly tested, validated as well as modified. In fact such are the dynamic prediction systems which the decision makers need and this is what the data mining technology can provide (Olson and Delen, 2008).

The IT infrastructure of any business at a given point of time depends on the following factors.

- The business requirements of the enterprise at that particular point of time
- The available technology at that point of time
- The accumulated investments of the enterprise from earlier technological inventions

The business requirements of the enterprise are constantly changing and are growing at a rapid pace. From the day to day clerical recording of business transactions business enterprises have now incorporated automated processes in place. Technology has increased the computing power as well as the communications capabilities to a considerable extent. In the world of today the IT professionals have a twofold responsibility. The first one is to meet the

information requirements through information technology and the second is to integrate the new technology into the existing enterprise architecture. In hindsight the IT professional must ensure that the IT enterprise infrastructure is vibrant and keeps on changing. At the same time the IT professionals must learn new buzzwords, technologies, evaluate the new tools as well as maintain ties with the technology partners (Nagabhushana, 2006)

In fact the evolution process has changed from one way to a two way interaction process. One of the prime reasons for such a change is the fact that data mining happens to be an emerging area of business. The needs of an organization calls for a need to integrate from a single process to multiple ones in which the interests of both the parties are of paramount importance. Many issues have been found in the related communities and these issues cannot be tackled simply by the internal techniques. This calls for the development of agent driven data mining as well as data mining driven agents (Cao, 2009).

4. IMPORTANCE OF DATA MINING TECHNOLOGIES

Data mining is a technology which is powerful in stature and helps the company to reflect on the information which they have collected in the form of data relating to consumer behaviour along with potential customers. Information within a data is revealed which cannot be discovered by reports as well.

The techniques of data mining are used in various areas and numerous commercial systems are available. The retail sector is one of the areas where data mining finds its prominence as it collects a large amount of information relating to customer sales, reports, consumption, and transportation along with service. The quantity of data collected also continues to increase with the ease and popularity of the business conducted on the web. Therefore the precise data mining technique should be identified from the host of available data mining techniques. The retail sector is the sector which sells goods and services directly to the customers in small quantities or an individual customer who will buy it and use for their own purposes. It is not only important in attracting new customers but also to ensure that the existing customers do not move away to the competitors. In the retail sector data mining can be used to understand the relationships and interactions among the organization as well as the customers to improve the marketing campaign by working on the outcomes to provide the customers with more focussed support as well as attention. The fact of the matter is that the cost of acquiring new customers is higher than retaining the old customers and this is the precise

reason on why companies focus on customer retention (Hruschka et al, 1998)

Data mining has emerged quickly emerged as a tool that will help the organizations to exploit the information assets. Data mining in the financial markets is all the more important as well. An organizational framework for the study of data mining is the urgent need of the hour. It is a challenging task owing to the multidisciplinary nature of the fast growing field (Han et al, 2006).

In fact great progress has been made in the domain of data mining in the last few years or so as several new process, techniques, systems along with applications have been developed.

Data mining plays a huge role in the telecommunications industry. Data mining is a technology that needs to coexist with the major other technologies in the organization. Data mining offers value across a wide spectrum of industries and can be used as a vehicle to increase profits by reducing costs or by raising revenue. Many organizations in the world of today use data mining to manage all phases of the customer life cycle.

This may include acquiring new customers, increasing the revenue from the existing customers as well as retaining the good customers. If the characteristics of the good customers are determined the company can target customers of similar nature. This in a way could be referred to as profiling as it also helps to ascertain the customers who are likely to leave the business. Some of the industries where data mining finds its prominence are

- Telecommunications as well as credit card companies are two of the leaders in applying data mining to detect the fraudulent use of their services
- Insurance companies as well as insurance companies are interested in data mining to reduce frauds
- Medical applications use data mining to predict the effectiveness of surgical procedures as well as medical tests.

5. DATA MINING TECHNOLOGIES IN IT COMPANIES

A number of companies attempted to change or redirect their efforts in the early part of the 1980's as well as the 1990's. In fact during this decade that customer service became a hot topic and everyone from the top to the low level of management was advised to take care of the customer in the best possible way. Companies that are forward looking focus on product development and believe

the fact that the customers are the life line of the business and they should be managed as well as maintained properly (Sumathi and Sivanandam, 2011).

The producers as well as the suppliers must put together the right blend of products for the ultimate customer. In fact the technology to develop these traits exists in the domain of data mining. Companies are able to take information from their own database and supplement with the additional information from the data compiler and then apply a predictive model to the augmented data using the data mining techniques. This will go a long way in understanding the vision of the customers in the days to come. It is companies that are in a stable environment which will feel the need for data mining technologies.

Data mining technologies will continue to mature and will develop in the days to come. In fact learning from data mining technologies will become available in the trade press as well as the commercial publications. They are driven by marketing as well as sales departments and are more popular in big companies with large databases at their disposal. The popularity of data mining techniques will coincide with the boom in data warehouse. It will also further underscore the importance of data quality in the domain of data warehouse implementations as well (Nagabhushana, 2006).

The credit card companies tend to notice the abnormal spending pattern of the customers and such patterns can expose the fraudulent use of the cards as well. Without data mining in place heaps of data lay under the carpet which if studied can determine new patterns as well as relationships.

Customers interact with the organization in a variety of ways (Ponniah, 2010). Interactions with the customers happen in three different ways during the customer life cycle which are

- Acquisition stage of the customer
- Value enhancement of the customer
- Retention of the customer

6. CONCLUSION

Data mining happens to be an active tool which is used by the government along with various organizations to establish trends relating to specific purposes. For example amazon.com uses data mining to promote the sales figure by selecting the pre sale items. On the other side of the coin there is various privacy issues related to data mining as already information is available and without realizing the

fact you tend to give out more of it. The key component in this regard is how you manage your personal information and what do you think is appropriate to give out.

Information is available through a variety of ways and each sort of information may be important to any one according to their needs. In this regard it is suggested to read the privacy policies of the various websites before giving out information. Data mining technologies have evolved leaps and bounds in the last decade or so with the online boom. Various software's have added a new dimension to this domain and the key is to utilize these techniques for one's own benefit as well as enhancing the profit of the organization. The key point of observation is data mining is here to stay big time

REFERENCES

1. Gopalan & Sivaselvan (2009). Data Mining. New Delhi: Ashoke
2. Sumathi, S. and Sivanandam (2011). Introduction to Data Mining and Its Applications. Berlin: Springer
3. Shmueli, G., Patel, N., and Bruce, P. (2010). Data Mining for Business Intelligence: Concepts, Techniques, and Applications. NJ: Wiley
4. Agarwal, B., and Tayal, S. (2009). Data Mining and Data Warehousing. New Delhi: Laxmi
5. Mena, J. (2007). Homeland Security Techniques and Technologies. Delhi: Laxmi
6. Skalak, S., Golden, T., and Clayton, M., and Pill, J. (2011). A Guide to Forensic Accounting Investigation. NJ: PWC
7. Olson, D., and Delen, D. (2008). Advanced Data Mining Techniques. Berlin: Springer.
8. Nagabhushana, S. (2006). Data Warehousing Olap And Data Mining. Delhi: New Age
9. Cao, L. (2009). n Data Mining and Multi-agent Integration. NY: Springer.
10. Hruschka, E., Watada, J., nad Nicoletti, M. (2011). Integrated Computing Technology: First International Conference, INTECH 2011.
11. Han, J., Kamber, M., nad Pei, J. (2011). Data Mining, Second Edition: Concepts and Techniques. SF: Morgan Kauffman
12. Ponniah, P. (2010). Data Warehousing Fundamentals for IT Professionals. NJ: Wiley