

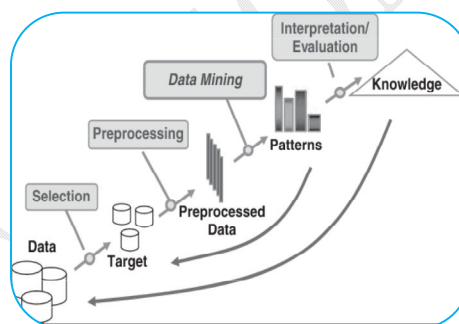


USEFULNESS OF DATA MINING IN KNOWLEDGE DISCOVER PROCESS

Prashant Devidas Jadhav
Research Scholar

ABSTRACT

Present era in an age often referred to as the information age. In this information age, because we believe that information leads to power and success, and thanks to sophisticated technologies such as computers, satellites, etc., we have been collecting tremendous amounts of information. Initially, with the advent of computers and means for mass digital storage, we started collecting and storing all sorts of data, counting on the power of computers to help sort through this amalgam of information. Unfortunately, these massive collections of data stored on disparate structures very rapidly became overwhelming. This initial chaos has led to the creation of structured databases and database management systems (DBMS). The efficient database management systems have been very important assets for management of a large corpus of data and especially for effective and efficient retrieval of particular information from a large collection whenever needed. The proliferation of database management systems has also contributed to recent massive gathering of all sorts of information. Today, we have far more information than we can handle: from business transactions and scientific data, to satellite pictures, text reports and military intelligence. Information retrieval is simply not enough anymore for decision-making. Confronted with huge collections of data, we have now created new needs to help us make better managerial choices. These needs are automatic summarization of data, extraction of the "essence" of information stored, and the discovery of patterns in raw data.



KEYWORDS: Women Dairy Farmer, Entrepreneur Behavior.

INTRODUCTION:-

Data mining is defined as the process of extracting information from a large set of data. In other words, we can say that data mining is mining from data. Terms involved in data mining such as knowledge search, query language, classification and prediction, decision tree induction, cluster analysis. Data mining is considered an

interdisciplinary field that incorporates computer science and statistical techniques. Note that the word "data mining" is a misnomer. It is primarily concerned with finding patterns and discrepancies between datasets, but not with the extraction of data itself. There is a huge amount of data available in the information industry. This data is of no use unless it is converted into useful

information. It is necessary to analyse this large amount of data and extract useful information from it.

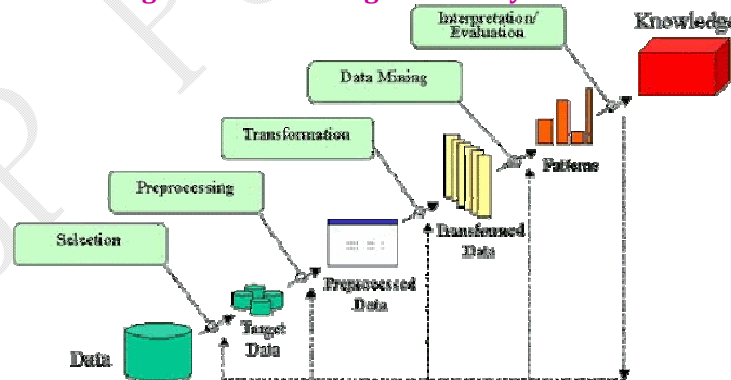
Extracting information from us is not just a process; Data mining also includes other processes such as data cleaning, data integration, data transformation, data mining, pattern evaluation and data presentation. Once all these processes are completed, we will use this information in many

applications such as fraud detection, market analysis, product control, science exploration, etc. Data mining also known as Knowledge Discovery (KDD) in a database or simply the Knowledge Discovery Process is a computerized process of opening up meaningful information or patterns that were not known before a set of data using methods such as database systems. Machine learning, artificial intelligence and statistics, Collaborative, predictive, clustering, and pattern identification are the objectives of DM's methods. It is run in a series of steps that are automatically or semi-automatically removed to find and retrieve hidden, unknown and interesting features from a large amount of data. Improved quality of the data obtained and perfectly functional methods are the two main aspects of data mining. In human society, data mining is applied in many areas, including the engineering field, customer relationship management, market analysis and weather forecasting. It can also be used in incorrect and incomplete data or information transactions.

In fact, DM is just one step in the overall process of KDD. KDD is a noncritical process for identifying valid, novel, perhaps useful, and fundamentally intelligible patterns of data. The process of KDD with DM component always depends on the algorithm used to obtain a set of samples from the given data. The DM process has various advantages compared to the existing traditional methods. For example, compared to traditional methods, the DM process requires fewer assumptions for the processing of regression and classification applications. DM and Knowledge Management allow the growth of domain expert decision by using additional experts and present both of them with competitive advantage. Managers and executives are users of the system at every level of control. DM serves as an interdisciplinary field with the basic objective of understanding results and finding data relationships, approving the benefits of automated techniques and tools, implementing sophisticated algorithms, discovering overlapping patterns, inconsistencies, structures and / or associations of data stored in a data warehouse or other data repository. In the manufacturing sector; Description and forecasting are the two initial objectives of DM. The behaviour of the model is predicted with predictive data mining, and the available databases are obtained to obtain existing data to generate future values of key variables, while interesting specimens with descriptive data mining are characterized by data. The boundary between the predictor and the descriptive DM is not sharp.

Most people consider data mining to be a “knowledge discovery in a database” or KDD equivalent process. Otherwise, it is also seen as a simple but necessary step in the search for knowledge in the database process. Being a process, the search for knowledge involves a repetitive process.

Figure 1.1 Knowledge Discovery Process



- Clear data (to remove noise or unrelated data),
- Data aggregation (in which multiple data sources are merged),
- Data selection (data related to analysis is extracted from a given database),
- Data Transformation (Mining is converted or aggregated into a suitable structure by summarizing data or by total operation),
- Data mining (a necessary step in implementing intelligent methods for data sampling),

- Sample evaluation (to identify really interesting patterns that show knowledge based on some interesting metrics) and
- Knowledge presentation (visual and knowledge representation techniques are applied to provide mining information to the end user).

Data mining is the application of an automated data analysis method to unravel past unknown relationships between data items. This usually involves the analysis of the data in the data warehouse. Data mining is widely useful for market analysis and management, corporate ud analysis, risk managements and fraud detecting domains.

What kind of information are we collecting? We have been collecting a myriad of data, from simple numerical measurements and text documents, to more complex information such as spatial data, multimedia channels, and hypertext documents. Here is a non-exclusive list of a variety of information collected in digital form in databases and in flat files.

Usefulness of Data Mining:

- **Business transactions:** Every transaction in the business industry is (often) “Memorized” for perpetuity. Such transactions are usually time related and can be inter-business deals such as purchases, exchanges, banking, stock, etc., or intra-business operations such as management of in-house wares and assets. Large department stores, for example, thanks to the widespread use of bar codes, store millions of transactions daily representing often terabytes of data. Storage space is not the major problem, as the price of hard disks is continuously dropping, but the effective use of the data in a reasonable time frame for competitive decision making is definitely the most important problem to solve for businesses that struggle to survive in a highly competitive world.
- **Scientific data:** Whether in a Swiss nuclear accelerator laboratory counting particles, in the Canadian forest studying readings from a grizzly bear radio collar, on a South Pole iceberg gathering data about oceanic activity, or in an American university investigating human psychology, our society is amassing colossal amounts of scientific data that need to be analysed. Unfortunately, we can capture and store more new data faster than we can analyse the old data already accumulated.
- **Medical and personal data:** From government census to personnel and customer files, very large collections of information are continuously gathered about individuals and groups. Governments, companies and organizations such as hospitals, are stockpiling very important quantities of personal data to help them manage human resources, better understand a market, or simply assist clientele. Regardless of the privacy issues this type of data often reveals, this information is collected, used and even shared, when correlated with other data this information can shed light on customer behaviour and the like.
- **Surveillance video and pictures:** With the amazing collapse of video camera prices, video cameras are becoming ubiquitous. Video tapes from surveillance cameras are usually recycled and thus the content is lost. However, there is a tendency today to store the tapes and even digitize them for future use and analysis.
- **Satellite sensing:** There are a countless number of satellites around the globe: some are geo-stationary above a region, and some are orbiting around the Earth, but all are sending a non-stop stream of data to the surface.
- **Games:** Our society is collecting a tremendous amount of data and statistics about games, players and athletes. From hockey scores, basketball passes and car-racing lapses, to swimming times, boxer’s pushes and chess positions, all the data are stored. Commentators and journalists are using this information for reporting, but trainers and athletes would want to exploit this data to improve performance and better understand opponents.
- **Digital media:** The proliferation of cheap scanners, desktop video cameras and digital cameras is one of the causes of the explosion in digital media repositories. In addition, many radio stations, television channels and film studios are digitizing their audio and video collections to improve the

management of their multimedia assets. Associations such as the NHL and the NBA have already started converting their huge game collection into digital forms.

- **CAD and Software engineering data:** There are a multitude of Computer Assisted Design (CAD) systems for architects to design buildings or engineers to conceive system components or circuits. These systems are generating a tremendous amount of data.
- **Virtual Worlds:** There are many applications making use of three-dimensional virtual spaces. These spaces and the objects they contain are described with special languages such as VRML. Ideally, these virtual spaces are described in such a way that they can share objects and places. There is a remarkable amount of virtual reality object and space repositories available. Management of these repositories as well as content-based search and retrieval from these repositories are still research issues, while the size of the collections continues to grow.
- **Text reports and memos (e-mail messages):** Most of the communications within and between companies or research organizations or even private people, are based on reports and memos in textual forms often exchanged by e-mail. These messages are regularly stored in digital form for future use and reference creating formidable digital libraries.

Data mining is the automatic or semi-automated analysis of large amounts of data to extract previously unknown interesting specimens. There are many big data mining techniques developed and used in data mining projects. These techniques include association, classification, clustering, forecasting, and sequential sampling. Obscure Logics is able to justify reasonably human reasoning by allowing partial subscriptions for data items in obscure subtext. The integration of obscure logic with data mining techniques has become a key component of machine learning to address the challenges posed by the large collection of natural data.

Since data mining is a computer-based technology, it is only natural to create principles of data mining using re-production neural networks. But the purpose of this study is to use data mining techniques to model behaviour patterns of human interaction with different systems. Human behaviour is always vague and it is very difficult to create patterns of human behaviour with crunchy neural network data mining systems. Modelling of erroneous and qualitative knowledge, as well as resolution of uncertainties at different stages is possible through the use of obscure sets.

CONCLUSION:

Few will compete with the potential of data mining tools to generate valuable business insights. However, like all technologies, the deployment of data mining must be driven by well-researched enterprise needs, as well as cost and usability considerations. Without specific effort, your mind is forming clusters and associations. When you see a man and a woman getting close to each other, you know they are either related or a couple. You see a woman coming out of a certain store and you immediately associate her with the image displayed in the store. Data mining systems make it easy for us to handle large amounts of data. Everything that is done in data mining can be done manually by humans but it will take much more time.

REFERENCES:

1. Anusha N. Rajashree and Srikant Bhat K. (2018), 'A Survey of Medical Data Mining Technique', International Journal of Science Engineering and Technology Research, Vol-7, Issue-1, pp. 11-17.
2. Behal Sunny and Krishan Kumar (2016), 'Trends in Validation of DDoS Research', International Conference on Computational Modeling and Security, Procedia Computer Science, 85 (2016), 7-15.
3. Chirag and Komal Sharma (2017), 'Data Mining Techniques in Healthcare: A Case Study', International Journal of Computer Science Trends and Technology, Vol-5, Issue-6, pp. 62-67.
4. Dhanalakshmi R. (2019), 'Efficient Disease Diagnosis System Using Data Mining Techniques', Submitted Thesis at Faculty of Information and Communication Engineering, Anna University, Chennai

5. Mohana S. (2016), 'A Novel Feature Selection Technique for Privacy Preservation in Data Mining Using Memetic Approach', Submitted Ph.D. Thesis at Faculty of Information and Communication Engineering Anna University, Chennai.
6. Neesha Jothi, Nur'Aini Abdul Rashid, Wahidah Husain (2015), 'Data Mining in Healthcare - A Review', Procedia Computer Science 72 (2015) 306 – 313.
7. Pareek Vikas and Rathi Megha (2018), 'Disease Prediction Model Using Data Mining Techniques', Submitted Ph.D. thesis at Department of Computer Science Banasthali Vidyapeeth, Rajasthan
8. Ramageri Bharati M. (2015), 'Data Mining Techniques and Applications', Indian Journal of Computer Science and Engineering, Vol-1, Issue-4, pp. 301-305.
9. Senthilkumar J.P. (2019), 'An Efficient Optimization Based Scheduling Mechanism for A Reliable MIMO-OFDM Communication System', Submitted Thesis at Faculty of Information and Communication Engineering Anna University, Chennai.
10. Thitimanan Damrongsakmethee and Victor-Emil Neagoe (2017), 'Data Mining and Machine Learning for Financial Analysis', Indian Journal of Science and Technology, Vol-10, Issue-39, pp. 1-7.