International Research Journal of Engineering and Management Studies (IRJEMS)
Volume: 03 Issue: 04 | April -2019 ISSN: 1847-9790 || p-ISSN: 2395-0126

Escalation of Data Mining-A Knowledge Expansion Phenomenon

Harpreet Kaur [1], Vinay Gautam [2]

[1] *Assistant Professor in Department of Computer Science and Applications*
[2] *Associate Professor Department Of Computer Science And Engineering*

***

Abstract - *Data mining is a multi-purpose discipline which includes knowledge discovery processes using databases technologies, machine learning, pattern recognition, artificial intelligence and data visualization .This paper describes briefly the definition of data mining, its functionality and various steps involved in knowledge management process. This paper also explains data mining techniques which have been developed to support knowledge management process. Finally, it explains the applications of data mining techniques in the process of knowledge management and related issues*

***Key Words*: Data mining, Knowledge management, Data Mining Techniques, Mining Applications and Issues.**

1. INTRODUCTION

Today there is a rapid growth of data, data collection and information availability. For this purpose different automated data collection tools, database systems are used. Major sources of plentiful data are business, web, e-commerce, transactions and stocks. Invention of various scientific technologies like remote sensing, bioinformatics, simulation requires appropriate data. So there is need for data mining, because "Necessity is mother of invention"-Data mining is automated analysis of massive data sets [1].
Data mining is derivation of useful patterns and information from a large amount of data. Once the patterns are found they can further be used to make certain decisions [2].It is a knowledge discovery process from huge amount of data means information extraction using different data analysis techniques. Extracted data is significant, previously unknown and potentially beneficial. Evolution of database technology is given below [1]:

- 1960's -Data Collection, Database Creation, IMS and network DBMS.
- 1970's- Relational Data Model, Relational DBMS Implementation.
- 1980's-RDBMS, Advanced Data Models, Application Oriented DBMS, Data Access, Structured Query Language, ODBC.
- 1990's- Data Mining, Data warehousing, multimedia databases, web databases, OLAP (On-line analytic processing).
- 2000's-Stream Data management and mining, Web Technologies and global information system.

Data mining is an essential part of knowledge management [4] [3].Knowledge management is a basis of data mining .It uses tools to extract useful knowledge from large datasets [4] [3]. Fig.1 shows knowledge discovery process in databases.

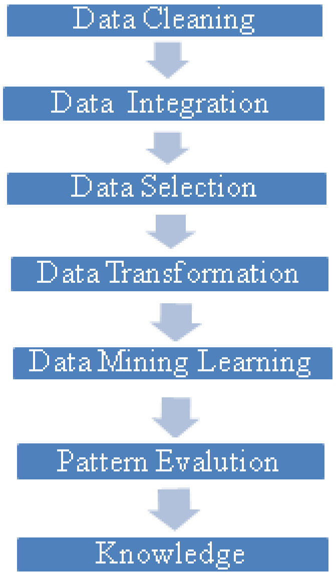

Fig.1 Data Mining Knowledge Discovery Process

Knowledge Process includes following steps[5]:

- Data Cleaning-To remove noise and incorrect data.
- Data Integration-To merge or integrate data from numerous sources.
- Data selection-To derive appropriate data for analysis.
- Data Transformation-To convert data into appropriate form for data mining.
- Data mining.
- Appraisal and estimation.

Basically Knowledge Process involves three Phases [2]:

- Exploration- Data is cleansed and converts into another form and important variables, then nature of data based on the problem are determined.
- Pattern Identification- Once the data is examined, purified and defined for the particular variables the next step is identification. Identify and select the patterns which make the best prediction.
- Deployment-Selected data are deployed for desired outcome.

1.1 DATA MINING TASKS/TECHNIQUES

There are many data mining techniques are used [8]:

Clustering

This data mining technique makes meaningful or useful cluster of objects that have similar characteristic. Clustering technique defines the classes and put objects in them. For example, identification of areas of similar land use in an earth observation database. Popular clustering techniques include K-means clustering and Expectation Maximization Clustering (EM) [16].

Classification

It is classic data mining technique based on machine learning. It is used to classify each item in a set of data into one of predefined set of categories. In classification there is a use of mathematical techniques. Major algorithms which are used in classification are Decision trees, Rule-based induction, neural networks, Memory (Case) based reasoning, Genetic algorithms, Bayesian networks etc [6]. For example people looking to lose weight have a few options like exercise, diet, weight loss pills, and surgery.

Regression

Regression is the oldest and most well-known statistical tool that is used in data mining. For the calculations of regression numerical values and mathematical formula requires. So in future, you simply put your new values, put them into the derived formula and get estimated values. But the drawback of regression is that it only works well with a quantitative data like weight, speed etc [16]. Starting from linear regression in which formula of a straight line is ($y = mx + b$) then determines the appropriate values for m and b to predict the value of y which depend upon a value of x. In advanced techniques like multiple regressions, it uses more than one input variable and allow for the fitting of more complicated models like quadratic equation.

Association rule learning

Association rules are if/then statements that help uncover relationships between seemingly unrelated data in a relational databases or other information repository [17]. An example of an association rule would be, if a customer frequently purchases same product simultaneously from mall, then use of this information for future to determine association. It helps to searches relationships between variables. These rules are useful for determining and predicting purchaser conduct, Association rules used in shopping basket data analysis, product clustering, catalog design etc. There are many association rules to build programs efficient of machine learning. Machine learning is a component of artificial intelligence (AI) that relates the establishment and study of systems that can learn from data.

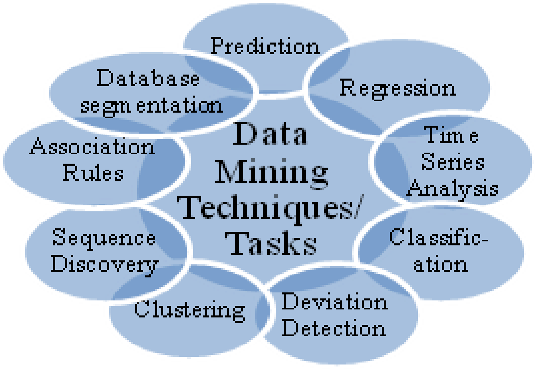

Fig.2 Classification of Data Mining Tasks

Prediction

This technique concludes relationship between independent variables and relationship between dependent and independent variables. For example prediction can be used to predict profit for the future, consider sale of cars as an independent variable, profit as a dependent variable. In next step taking previous sale and profit data, we can produce a fitted regression curve. This curve is used for company profit prediction. Prediction includes analyzing trends, classification, pattern matching, analyzing previous work company can make a prediction about future. This Technique links the estimate of future prospect. Using data mining technology straight, you can create a model or use a model created by someone else. Same data used for training and testing in predictive analytics[6].

Sequential Patterns

Sequential patterns analysis is data mining technique which attempt to find similar patterns in data transaction over a business period. The detected patterns are used for further business analysis and identify relationships among data. So that presence of one set of patterns is proceed by another set of patterns in a database over a period of time . For Example to understand long term customer by examining their buying habits.

Deviation Detection

Relatively new available data mining tool used marketable. It is a source of true discovery because it identifies an extreme deviation from the mean, which express deviation from some previously known assumptions and standards. It can be performed using statistics and visualization techniques .For example fraud detection application in the use of credit cards and insurance claims, quality control etc [17].

Similar Time Sequence Discovery

This technique searches links between two sets of data that are time-dependent .It is based on the degree of similarity between the patterns which are both time series display. For example new home purchaser will purchase goods such as cookers, freezers, and washing machines [17].

Database Segmentation

Segmentation partition a database into an unknown number of segments, or clusters, of similar records. It discovers homogeneous sub-populations in a database to improve the accuracy of the profiles. It uses neural clustering techniques which allow data inputs and methods used to calculate distance between records and further analysis. This approach is less precise than other techniques but less sensitive to duplicate and irrelevant features .Example of database segmentation involve customer profiling and direct marketing etc. [17].

2 ASSEMBLAGE OF THREE TECHNOLOGIES

○ **Increasing Computing Power**

Computing power become doubles every 18 months in Moore's law and many powerful workstations exist and parallel processing enhances Cost effective servers (SMPs).

○ **Enhanced Data Collection**

In Data mining there are four steps first data collect- ion next data access, navigation and final data mining. Here for the better use appropriate amount of data required [16].

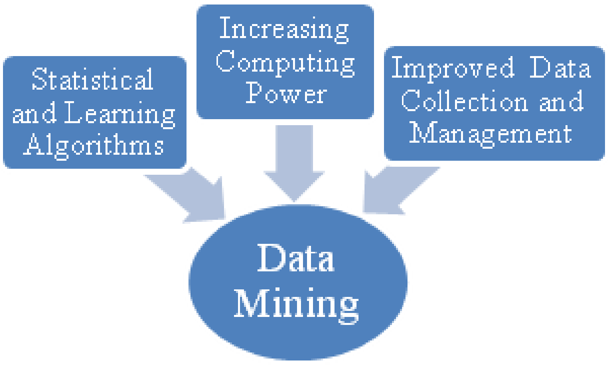
Statistical and Learning Algorithms
Increasing Computing Power
Improved Data Collection and Management
Data Mining

Fig.3 Conjunction of Technologies in Data Mining

- **Improved Algorithms**

Implementing various mining techniques improves computing technology and succeeds in reaching manual data mining. For this purpose machine learning is a good example. A lot of data mining research focused on improving existing techniques to get small percentage benefits.

2.1 DATA MINING APPLICATIONS

There are two types of applications of data mining. First one is generic and second is domain specific [6].The generic application is need to an intelligent system. Because using these systems certain decision like selection of data, selection of data mining method, presentation and interpretation of the result done .[4][7].

A. In data mining classification technique is used to classify Many type of **n**etwork **t**raffic.

B. Data Mining play very important role in medical science For example patient profiling and history generation, Diagnosis etc.

C. In any organization or company data mining provide Knowledge like store layout strategies, stock and Promotions and improve them. For this *Performing* **B**asket Analysis is used.

D. Data mining can be used *by* **S**ports **O**rganizations for the Statistical analysis, pattern discovery, as well as outcome Predictions, because large amount of data for each player is available.

E. The data mining algorithms are used that can properly Account for the temporal nature of the data and the Character of group interaction using identifies patterns.

F. Data mining methods are also used in **w**eb based **e**ducation. This knowledge is very useful for the teacher or any author, who could decide what changes will be the Most appropriate to improve the course Effectiveness.

G. In agriculture there are various methods like pattern Recognition, sensing and actuator technologies as well Signal processing which is provided by data mining. New Technologies like wireless networks, new environmental Sensors, robots introduced.

H. In National Security Agency/Governments [16] Data Mining helps government agency by cultivating and Analyzing records of financial transaction to build patterns that can detect criminal activities.

I. In manufacturing using data mining in operational Data, manufacturers can detect faulty equipments and Determine effective control parameters [16].

2.2 ISSUES/DISADVANTAGES OF DATA MINING

But still there are many pending issues related to it [9] [11]:

Privacy Issues

Due to enhancement of social networks, forums blogs there are several privacy issues, people are afraid to give their information because it potentially causing them a lot of trouble. Organizations collect information about their customers in many ways for understanding their purchasing behaviors trends etc [16]. Suppose company may be acquired by other or closed then personal information they own probably is sold to other or leak.

Security issues

Personal information about employee and customers including social security number, birthday, payroll etc is a big security issues. There have been a lot of cases that hackers were accesses and stole data of customers and employee .So the credit card stolen and identity theft become a big problem.

Information exploitation/inaccurate information

Information collected for ethical purposes through data mining can be misused. This information is exploited by immoral people. All data mining technique is not perfectly accurate. So if wrong information is used for decision-making the outcome will also be incorrect [16].

Scalability Problem

In data mining dealing with large data sets arise the issues of performance and efficiency in data mining methods because different sources of knowledge like pattern evaluation ,web , data streaming etc are major Issues[1].

3. CONCLUSION

In any organization, knowledge is a major capital and the management of knowledge assets has become a strong demand for development. Deriving the useful information for decision making is essential part. So data mining is a main part of knowledge management. It is a progressive gain from transforming data into information and information into knowledge. Various pattern recognition technologies, statistical and mathematical techniques of data mining helps analysts to recognize significant facts, relationships, discovery of knowledge, trends, patterns, exceptions and different problem also. There are many steps involves in it like selection of specific data for data mining, cleaning is next step and then transformation of data, In next step separation of information patterns for knowledge generation and in final step knowledge generation by interpretation of the information patterns . Data mining is used in medical science, sports organizations, in agriculture, web applications, identifying patterns, sales forecasting, performing basket analysis etc. But still pending issues like security and social issues, privacy issues user interface issues, performance issues etc. Now a day's Data Mining plays very vital role in the field of computer science. It will continue and even increase over coming decades.

Table 1

Figure	Name of Figure
Fig.1	Fig.1: Data Mining Knowledge Discovery Process
Fig.2	Fig.2: Classification of Data Mining Tasks
Fig.3	Fig.3:Conjuction of Technology in Data Mining

REFERENCES

[1] (IS698) Min Song, Data Mining. [Online].Available: http://web.njit.edu /~song/courses/is698/week11.ppt.
[2] Bahrain M. Ramageri,"Data Mining Techniques and Applications."*Indian Journal of Computer Science and Engineering. Vol. 1 No. 4 301-305*.
[3] S. Tipawa, T. Kulthida,"Data Mining and Its Applications for Knowledge Management: A Literature Review from 2007 to 2012,"*International Journal of Data Mining & Knowledge Management Process (IJDKP) Vol.2, No.5, September 2012.*
[4] J. Dawei (2011)," The Application of Date Mining in Knowledge Management.2011 International Conference on Management of e- Commerce and e-Government," *IEEE Computer Society, 7-9. doi: 10.1109/ICMeCG.2011.58.*
[5] Bernard Chen (2010) Chapter 1 Introduction to Data Mining. [Online]
[6] Rupali, G. Gaurav," Data Mining: Techniques, Applications and Issues," *International Journal of Advanced Research in Computer Science and Electronics Engineering (IJARCSEE) Volume 2, Issue 2, February 2013.*
[7] Dunham, M. H., Sridhar S., "Data Mining: Introductory and Advanced Topics," *Pearson Education, New Delhi, ISBN: 81-7758-785-4, 1st Edition, 2006*
[8] Fayyad, U., Piatetsky-Shapiro, G., and Smyth P., "From Data Mining to Knowledge Discovery in Databases, *"I Magazine, American Association for Artificial Intelligence, 1996*.
[9] Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinartz, T., Shearer, C. and Wirth, R. "CRISP-DM 1.0 : Step-by-step data mining guide, NCR Systems Engineering Copenhagen (USA and Denmark), DaimlerChrysler AG (Germany), SPSS Inc. (USA) and OHRA Verzekeringenen Bank Group B.V (The Netherlands), 2000".
[10] Data Mining Techniques.[Online].Available: http://www.zentut.com/data-mining/data-mining-techniques/
[11] Data Mining Issues. [Online].Available: http://www.seas.gwu.edu/~bell/csci243/lectures/introduction.pdf
[12] SlobodanVucetic (2004) CIS527: Data Warehousing, Filtering, and Mining [Online].Available:" http://www.ist.temple.edu/ ~vucetic/ cis526 fall2004 / lecture3.ppt"
[13] S. P. Deshpande1 and Dr. V. M. Thakare2, "Data Mining System and Applications: A Review," *IJDPS, Vol.1, No.1, September 2010, pp.36-41.*
[14] P. Neelamadhab, M. Pragnyaba and P. Rasmita, "The Survey of Data Mining Applications and Feature Scope," *IJCSEIT, Vol.2, No.3, June2012, pp.48-53.*
[15] M. Munk, J. Kap usta, and P. Svec,"Data reprocessing valuation for web log mining: reconstruction of activities of web visitor,*" Procedia CS, 1(1):2273-2280, 2010.*
[16] S. Hemlata, S. Shalini and G. Seema," A Brief Overview on Data Mining Survey," *International Journal of Computer Technology and Electronics Engineering (IJCTEE) Volume 1, Issue 3.*
[17] Data Mining Transparencies. [Online].Available: http://www.slideshare.net/ManjuSingla/ch35-16973414