Volume 9 • Issue 1 Sept 2017 - March 2018 pp. 53-55 available online at www.csjournals.com

Data Mining Techniques

Sunny Sharma[1], Arjun Kumar[2]
[1]Research Scholar1, Department of CSE, Arni University kathgarh, Indora, HP.
[2]Department of CE, Arni University kathgarh, Indora, HP
sunny202658@gmail.com, arjun.devq @gmail.com

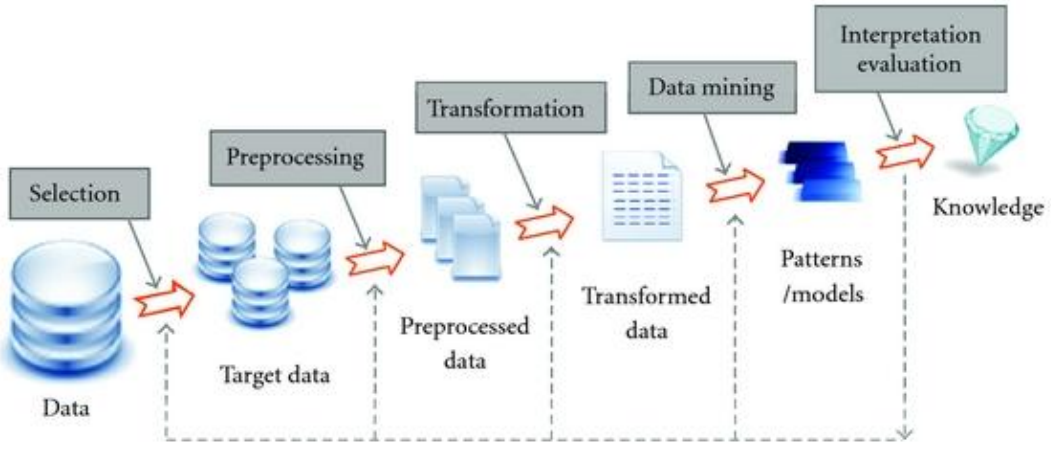
Abstract: The part of KDD dealing with the analysis of the data has been termed as data mining. It is a procedure which finds useful patterns from large amount of data. This data mining definition has business flavor and for business environments. However, data mining is a process that can be applied to any type of data ranging from weather forecasting, electric load prediction, product design, etc. This paper discuses the requirements of DM, and gives the overview of DM techniques such as Association, Classification, Clustering, Prediction, decision tree approach.
Keywords: Data Mining, Knowledge Discovery, Clustering, Association.

1. INTRODUCTION

The development of information technology has generated large amount of databases and huge data in various areas. To discover the information or knowledge from the abundant amount of data we have to call KDD (knowledge discovery in databases). Data Mining is a part [1] of KDD that deals with the analysis of data. The research in databases and information technology has given rise to an approach to store and manipulate this precious data for further decision making. Data mining is a process of extraction of useful information and patterns from huge data. It is also called as knowledge discovery process, knowledge mining from data, knowledge extraction or data /pattern analysis. Data warehousing [2] is one of the most important research areas related to DM. A data warehouse is a read-only database developed for analyzing business situations and supporting decision makers. The data warehouse includes large volumes of subject- oriented data, where all levels of an organization can find the information in a timely manner. DM goes together with the data warehousing which is necessary to organize historical information gathered from large-scale client/server-based applications.

2. KNOWLEDGE DISCOVERY PROCESS

Knowledge discovery is a process that extracts implicit, potentially useful information or knowledge from the data.



The knowledge discovery process is described as follows:
Let's examine the knowledge discovery process in the diagram above in details:

- Data comes from variety of sources is integrated into a single data store called target data

- Data then is pre-processed and transformed into standard format.
- The data mining algorithms process the data to the output in form of patterns or rules.
- Then those patterns and rules are interpreted to new or useful knowledge or information.

The ultimate goal of knowledge discovery and data mining process is to find the patterns that are hidden among the huge sets of data and interpret them to useful knowledge and information. As described in process diagram above, data mining is a central part of knowledge discovery process.

3. DATA MINING TECHNIQUES

There are several major data mining techniques have been developed and used in many data mining projects recently including association, classification, clustering, prediction, statistical analysis and sequential patterns etc., are used for knowledge discovery from databases. The techniques are studied from the [3].

Association:

Association is one of the best known data mining technique. In association, a pattern is discovered based on a relationship of a particular item on other items in the same transaction. For example, the association technique [4] is used in market basket analysis to identify what products that customers frequently purchase together. Based on this data businesses can have corresponding marketing campaign to sell more products to make more profit.

Applications: market basket data analysis, cross-marketing, catalog design, loss-leader analysis, etc.

Classification

Classification is a classic data mining technique based on machine learning. Basically classification is used to classify each item in a set of data into one of predefined set of classes or groups. For Example, Teachers classify students' grades as A, B, C, D, or F. Classification method [3] makes use of mathematical techniques such as decision trees, linear programming, neural network and statistics.

In classification, we make the software that can learn how to classify the data items into groups. For example, we can apply classification in application that "given all past records of employees who left the company, predict which current employees are probably [3] to leave in the future." In this case, we divide the employee's records into two groups that are "leave" and "stay". And then we can ask our data mining software to classify the employees into each group.

Classification Techniques

- Regression
- Distance
- Decision Trees
- Rules
- Neural Networks

Clustering

Clustering is "the process of organizing objects into groups whose members are similar in some way". A cluster is therefore a collection of objects which are "similar" between them and are "dissimilar" to the objects belonging to other clusters. We can take library as an example. In a library, books have a wide range of topics available. The challenge is how to keep those books in a way that readers can take several books in a specific topic without irritate. By using clustering technique [5], we can keep books that have some kind of similarities in one cluster or one shelf and label it with a meaningful name. If readers want to grab books in a topic, he or she would only go to that shelf instead of looking the whole in the whole library.

Clustering Techniques

- Artificial neural network (ANN)
- Nearest neighbor search
- Neighbourhood components analysis
- Latent class analysis
- Affinity propagation

Prediction

The prediction as it name implied is one of a data mining techniques that discovers relationship between independent variables and relationship between dependent and independent variables. In data mining independent variables are attributes already known and response variables are what we want to predict unfortunately, many real-world problems are not simply prediction [7] For instance, sales volumes, stock prices, and product failure rates are

all very difficult to predict because they may depend on complex interactions of multiple predictor variables. Therefore, more complex techniques (e.g., decision trees) may be necessary to forecast future values.
For instance, prediction analysis technique can be used in sale to predict profit for the future if we consider sale is an independent variable, profit could be a dependent variable. Then based on the historical sale and profit data, we can draw a fitted regression curve that is used for profit prediction.

Sequential Patterns
Sequential patterns analysis in one of data mining technique that seeks to discover similar patterns in data transaction over a business period. The uncover patterns are used for further business analysis to recognize relationships [4] [6] among data.

4. CONCLUSION

Data mining is a "decision support" part of KDD which we search for patterns of information in data. In this paper, we explain data mining and the techniques used in discovering knowledge from the collected data. Data mining techniques such as classification, clustering, prediction, association and sequential patterns etc it helps in finding the patterns to decide upon the future trends in businesses to grow. Data Mining Techniques thoroughly acquaints you with the new generation of data mining that's why data mining is considered one of the most important frontiers in database and information systems and one of the most promising interdisciplinary developments in Information Technology also.

References:

1. Zhao, Y. (2015). Data mining techniques.
2. Inmon, W. H. (2005). Building the data warehouse. John wiley & sons.
3. Raval, K. M. (2012). Data Mining Techniques. International Journal of Advanced Research in Computer Science and Software Engineering, 2(10).
4. Ngai, E. W., Xiu, L., & Chau, D. C. (2009). Application of data mining techniques in customer relationship management: A literature review and classification. Expert systems with applications, 36(2), 2592-2602.
5. Berkhin, P. (2006). A survey of clustering data mining techniques. Grouping multidimensional data, 25, 71.
6. Srikant, R., & Agrawal, R. (1996). Mining sequential patterns: Generalizations and performance improvements. Advances in Database Technology—EDBT'96, 1-17.
7. Sharma, S., & Rana, V. (2017). Web Personalization through Semantic Annotation System. Advances in Computational Sciences and Technology, 10(6), 1683-1690.