

The Need Of Data Mining In Web Services: A Research Study

Sanjay Sudhakarrao Kulkarni

Research Scholar, JJT University, Rajasthan, India, Kulsan72@gmail.com

Abstract:

Data mining is a process which finds useful patterns from large amount of data The process of extracting previously unknown, comprehensible and actionable information from large databases and using it to make crucial business decisions - Simoudis 1996 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. Data mining also can be defined as the computer-aid process that digs and analyzes enormous sets of data and then extracting the knowledge.

Keywords: Knowledge discovery is a process, Data mining Techniques.

1 Introduction

The development of information technology has generated large amount of databases and huge data in various areas. 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.

Briefly speaking, data mining refers to extracting useful information from vast amounts of data. Many other terms are being used to interpret data mining, such as knowledge mining from databases, knowledge extraction, data analysis, and data archaeology. Nowadays, it is commonly agreed that data mining is an essential step in the process of knowledge discovery in databases, or KDD. In this paper, based on a broad view of data mining functionality, data mining is the process of discovering interesting knowledge from large amounts of data stored either in databases, data warehouses, or other information repositories.

2. Knowledge Discovery Process



Email id's:- aiirjpramod@gmail.com, pramodedu@gmail.com | website :- www.aiirjournal.com Chief Editor:- Pramod P.Tandale | Mob. No.09922455749

3 Need Of Data Mining

Data mining is to extract information from large amount of a data base. There are two main reasons to use data mining as a rapidly increase demands of data. These are: Too much data and too little information. There is a need to extract useful information from the data and to interpret the data. Existing infrastructure.

4 The Past Introduction Of Data Mining

The term "Data mining" was announced in the 1990s, but data mining is the development of a field with a long times gone by Data mining roots are traced back along three family lines: classical statistics, artificial intelligence, and machine learning:

Statistics are the foundation of most technologies on which data mining is built, e.g. regression analysis, standard distribution, standard deviation, standard variance, discriminate analysis, cluster analysis, and confidence intervals. All of these are used to study data and data relationships.

Artificial intelligence, or AI, which is built upon heuristics as opposed to statistics, attempts to apply human-thought-like processing to statistical problems. Certain AI concepts which were adopted by some high-end commercial products, such as query optimization modules for Relational Database Management Systems (RDBMS).

Machine learning is the union of statistics and AI. It could be considered an evolution of AI, because it blends AI heuristics with advanced statistical analysis. Machine learning attempts to let computer programs learn about the data they study, such that programs make different decisions based on the qualities of the studied data, using statistics for fundamental concepts, and adding more advanced AI heuristics and algorithms to achieve its goals.

Data mining, in many ways, is fundamentally the adaptation of machine learning techniques to business applications. Data mining is best described as the union of historical and recent developments in statistics, AI, and machine learning. These techniques are then used together to study data and find previously-hidden trends or patterns within.

Necessity is the mother of invention. Since ancient times, our ancestors have been searching for useful information from data by hand. However, with the rapidly increasing, in the 1950s, volume of data in modern times, more automatic and effective mining approaches are required. Early methods such as Bayes' theorem in the 1700s and regression analysis in the 1800s were some of the first techniques used to identify patterns in data. After the 1900s, with the proliferation, ubiquity, and continuously developing power of computer technology, data collection and data storage were remarkably enlarged. As data sets have grown in size and complexity, direct hands-on data analysis has increasingly been augmented with indirect, automatic data processing. This has been aided by other discoveries in computer science, such as neural networks, clustering genetic algorithms Decision trees in the 1960s and support vector machines in the 1980s.

Data mining is the process of applying these methods to data with the intention of uncovering hidden patterns [3]. Data mining or data mining technology has been used for many years by many fields such as businesses, scientists and governments. It is used to sift through volumes of data such as airline passenger trip information, population data and marketing data to generate market research reports, although that reporting is sometimes not considered to be data mining.

Data mining commonly involves four classes of tasks [1]:

classification, arranges the data into predefined groups;

Aayushi International Interdisciplinary Research Journal (Indexed With IIJIF and IJIndex) Vol - III **Issue-VIII** AUGUST 2016 **ISSN 2349-638x Impact Factor 2.147**

- clustering, is like classification but the groups are not predefined, so the algorithm will try to group similar items together;
- regression, attempting to find a function which models the data with the least error; and
- association rule learning, searching for relationships between variables.

According to Han and Kamber [2], data mining functionalities include data characterization, data discrimination, association analysis, classification, clustering, outlier analysis, and data evolution analysis. Data characterization is a summarization of the general characteristics or features of a target class of data. Data discrimination is a comparison of the general features of target class objects with the general features of objects from one or a set of contrasting classes. Association analysis is the discovery of association rules showing attribute-value conditions that occur frequently together in a given set of data. Classification is the process of finding a set of models or functions that describe and distinguish data classes or concepts, for the purpose of being able to use the model to predict the class of objects whose class label is unknown. Clustering analyzes data objects without consulting a known class model. Outlier and data evolution analysis describe and model regularities or trends for objects whose behavior changes over time.

5 Performance Of Data Mining

There are several major data mining techniques have been developed and used in data mining projects recently including association, classification, clustering, prediction and sequential patterns etc., are used for knowledge discovery from databases.



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 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.

Aayushi International Interdisciplinary Research Journal (Indexed With IIJIF and IJIndex) Vol - III **Issue-VIII** AUGUST 2016 **ISSN 2349-638x Impact Factor 2.147**

Types Of Association Rules: Different types of association rules based on

- * Types of values handled
- * Boolean association rules
- * Quantitative association rules
- * Levels of abstraction involved
- * Single-level association rules
- * Multilevel association rules
- Dimensions of data involved **
- erdisciplinary Ros * Single-dimensional association rules
- ٠ Multidimensional association rules

Classification

Goal: Provide an overview of the classification problem and introduce some of the basic algorithms. 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 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 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.



<u>6 Classification Techniques</u>

- Regression \Leftrightarrow
- * Distance
- ÷ **Decision Trees**

- Rules
- Neural Networks

Clustering

Clustering is "the process of organizing objects into groups whose members are similar in some way". 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, 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.

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 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 among data.

7 Conclusion

Data mining is a "decision support" process in which we search for patterns of information in data. In other words, Data mining has importance regarding finding the patterns, forecasting, discovery of knowledge etc in different business domains. 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 has wide application field almost in every industry where the data is generated 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. Data mining offers promising ways to uncover hidden patterns within large amounts of data. These hidden patterns can potentially be used to predict future behavior. The availability of new data mining algorithms, however, should be met with caution. First of all, these techniques are only as good as the data that has been collected. Good data is the first requirement for good data exploration. Assuming good data is available, the next step is to choose the most appropriate technique to mine the data. However, there are tradeoffs to consider when

Aayushi International Interdisciplinary Research Journal (Indexed With IIJIF and IJIndex)Vol - IIIIssue-VIIIAUGUST2016ISSN 2349-638xImpact Factor 2.147

choosing the appropriate data mining technique to be used in a certain application. There are definite differences in the types of problems that are conductive to each technique. The "best" model is often found by trial and error: trying different technologies and algorithms. Often times, the data analyst should compare or even combine available techniques in order to obtain the best possible results.

References

- [1] Gorunescu, F, Data Mining: Concepts, Models, and Techniques, Springer, 2011.
- [2] Han, J., and Kamber, M., Data mining: Concepts and techniques, Morgan-Kaufman Series of Data Management Systems San Diego:Academic Press, 2001.
- [3] Neelamadhab Padhy, Dr. Pragnyaban Mishra and Rasmita Panigrahi, "The Survey of Data Mining Applications and Feature Scope, International Journal of Computer Science, Engineering and Information Technology (IJCSEIT)", vol.2, no.3, June
- [4] Fayadd, U., Piatesky -Shapiro, G., and Smyth, P. From Data Mining To Knowledge Discovery in Databases", The MIT Press, ISBN 0–26256097–6, Fayap, 1996.
- [5] Jing He, Advances in Data Mining: History and Future, Third international Symposium on Information Technology Application, 978-0-7695- 3859-4, IEEE, 2009.
- [6] Berry, M.and Linoff, G., Master Data Mining: The Art and Science of Customer Relationship Management, Wiley publisher, 2000.
- [7] Fayyad, U.M., Piatsky Shapiro, G. and Smyth, P. (1996) From Data Mining to Knowledge Discovery in Data Base, AI Magazine, pp.37-54
- [8] Feelders, A., Daniels, H. and Holsheimer, M. (2000) 'Methodological and Practical Aspects of Data Mining', Information and Management, pp.271-281.
- [9] Hampshire, D. A. and Rosborough, B. J. (1993) 'The evolution of decision support in a manged care organization', Topics in healthcare financing, 20, 2, pp26-37
- [10] Prins, S and Stegwee, R. A. (2000) 'Zorgproducten geintegreerde infromatiesystemen', (in Dutch) Handboek sturen met zorgproducten, F3100-3, december.
- [11] Sheng, O. R. Liu (2000), 'Decision support for healthcare in a new information age', Decision Support Systems, 30, pp101-103.
- [12] Dutta, A. and Heda, S. (2000) 'Information systems architecture to support managed care business process', Decision Support Systems, 30, pp 217- 225.
- [13]Dr. Lokanatha C. Reddy, A Review on Data mining from Past to the Future, International Journal of Computer Applications (0975 – 8887) Volume 15– No.7, February 2011
- [14] Usama Fayyad, Gregory Piatetsky-Shapiro, and Padhraic Smyth, From Data Mining to Knowledge Discovery in Databases, Al Magazine Volume 17 Number 3 (1996)
- [15] http://www.slideshare.net/Annie05/sequential-pattern-discovery-presentation
- [16]http://dataminingtools.net/wiki/introductiontodatamining.php
- [17] http://www.dataminingtechniques.net
- [18]http://www.slideshare.net/huongcokho/data-mining-concepts