ABSTRACT:

Data mining is the practice of examining large pre-existing databases in order to generate new information. The paper discusses few of the data mining techniques, applications and some of the organizations which have adapted data mining technology to improve their businesses and found excellent results. Data mining on large databases have been a very big concern in research community due to the difficulty of analyzing large volumes of data using only traditional OLAP tools. This sort of process takes up a lot of computational power, memory and disk input and output, which can only be provided by parallel computers. Thus here we present to you a brief discussion on how the database technology can be integrated to data mining techniques. We also learn about the life cycle of data mining.

Keywords: Data mining, warehousing, life cycle, applications.

INTRODUCTION

Data mining, the extraction of hidden information from large databases, is one of the powerful technologies which has a great potential to help other companies focus on the most important information in their data warehouses. The concept of data mining tools predict the future trends and behaviors which allows businesses to make knowledge driven and clever decisions. Data mining techniques have increasingly been studied a lot, especially because of their application in real world databases. The only big problem that arises is that the databases tend to be very large and these topics repeatedly scan the entire set. Even though the term data mining was introduced in the 90’s, the concept of data mining goes back to many years. Data mining has reached its current form after going through many phases of research and studies. what we have today here is a competitive world of information and hence all are trying to make the best use of their data to make business profitable and successful. Storage and collection of data on computers, tapes and disks started in the 60’s. next major and evolutionary step in data mining happened during the 80’s with the introduction of relational databases and structured query languages.
This sure helped the users to know more about the data stored in relational databases using structured query languages. The next main introduction was about data warehousing which happened during the 90’s. Online analytic processing and multidimensional databases contributed to the growth of data warehousing. But now, data mining is the emerging technique. If we analyze each step of evolution of data mining, we get to observe clearly that each step is built upon the previous step. In addition to large data warehouses and basic statistics, a major part of data mining is artificial intelligence, commonly known as AI. This artificial intelligence was started in the 80’s with a set of algorithms which were designed to teach a computer how to “learn” by itself. As they developed, these algorithms became valuable data manipulation tools and hence they were applied to a large sets of data. The data mining software, combined with the artificial intelligence technology was able to generate its own relationship between the data. AI gave way to machine learning. It is defined as “the ability of a machine to improve its performance based on previous results.”

DATA WAREHOUSING

The term data warehouse was first coined by Bill Inmon in 1990. According to Bill, a subject oriented, integrated, non volatile, time variant collection of data. This data plays a major role for analysts to take informed decisions in an organization. Everyday, an operational database undergoes frequent changes on account of the transactions that take place. If a business executive wants to analyze any previous feedback on any data such as a supplier, product, or any consumer data, then the executive will not have any data available to analyze because all the previous data has been updated due to transactions. A data warehouse provides us some generalized and consolidated data in multidimensional view. It also provides us Online Analytical Processing Tools (OLAP). These tools help us in interactive and effective data analysis in a multidimensional space. This type of analysis will result in data generalization and data mining. Data mining functions classification, clustering, association, prediction can all be integrated using OLPA operations to enhance the interactive mining of knowledge at multiple level abstraction. This is why data warehousing has become a very important platform for data analysis and online analytical processing.
DATA MINING LIFE CYCLE

The life cycle of data mining project consists of six phases[2,4]. The sequence of the phases is not rigid, moving back forth between phases is always required. It depends on the outcome of each phase. The main phases are:

Business Understanding:
This phase focuses on understanding the project objectives and requirements from a business perspective, then converting this knowledge into a data mining problem definition and a preliminary plan designed to achieve the objectives.

Data Understanding:
It starts with an initial data collection, to get familiar with the data, to identify data quality problems, to discover first insights into the data or to detect interesting subsets to form hypothesis for hidden information.

Data Preparation:
In this stage, it collects all the different data sets and construct the varieties of activities basing on the internal raw data.

Modeling:
In this phase, various modeling techniques are selected and applied and their parameters are calibrated to optimal values.

Deployment:
The purpose of the model is to increase knowledge of the data, the knowledge gained will need to be organized and presented in a way that the customer can use it. The deployment phase can be as
simple as generating a report or as complex as implementing a repeatable data mining process across the enterprise.
Visualizing Data Mining Model

Evaluation:

In this stage the model is thoroughly evaluated and reviewed. The steps executed to construct the model to be certain it properly achieves the business objectives. At the end of this phase, a decision on the use of the data mining results should be reached.
The main objective of data visualization is the overall idea about the data mining model. In data mining most of the times we are retrieving the data from the repositories which are in the hidden form. This is the difficult task for a user. So this visualization of the data mining model helps us to provide utmost levels of understanding and what the data mining process has discovered, it is a much bigger leap to take the output of the system and translate it into an actionable solution to a business problem. The data mining models are of two types Predictive and Descriptive.
The predictive model makes prediction about unknown data values by using the known values. Ex. Classification, Regression, Time series analysis, Prediction etc. The descriptive model identifies the patterns or relationships in data and explores the properties of the data examined. Ex. Clustering, Summarization, Association rule, Sequence discovery etc.
The term clustering means analyzes the different data objects without consulting a known class levels. It is also referred to as unsupervised learning or segmentation. It is the partitioning or segmentation of the data into groups or clusters. The clusters are defined by studying the behavior of the data by the domain experts. The term segmentation is used in very specific context; it is a process of partitioning of database into disjoint grouping of similar tuples. The association rule finds the association between the different attributes.

Data Mining Applications

In this section, we have focused some of the applications of data mining and its techniques analyzed respectively.

Data mining is used for market basket analysis

Data mining technique is used in MBA (Market Basket Analysis). When the customer want to buying some products then this technique helps us finding the associations between different items that the customer put in their shopping buckets. Here the discovery of such associations that promotes the business technique. In this way the retailers uses the data mining technique so that they can identify that which customers intension (buying the different pattern). In this way this technique is used for profits of the business and also helps to purchase the related items.

Data Mining Applications can be generic or domain specific.

Data mining system can be applied for generic or domain specific. Some generic data mining applications cannot take its own these decisions but guide users for selection of data, selection of data mining method and for the interpretation of the results. The multi agent based data mining application [8, 10] has capability of automatic selection item sets. Sequence discovery is the process of finding new sequences in data.
Data Mining methods are used in the Web Education

Data mining methods are used in the web Education which is used to improve courseware. The relationships are discovered among the usage data picked up during students’ sessions. This knowledge is very useful for the teacher or the author of the course, who could decide what modifications will be the most appropriate to improve the effectiveness of the course. In the 21st century the beginners are using the data mining techniques which is one of the best learning method in this era. This makes it possible to increase the awareness of learners. Web Education which will rapidly growth in the application of data mining methods to educational chats which is both feasible and can be improvement in learning environments in the 21st century.

Conclusion

In this paper we briefly reviewed the various data mining applications. This review would be helpful to researchers to focus on the various issues of data mining. In future course, we will review the various classification algorithms and significance of evolutionary computing (genetic programming) approach in designing of efficient classification algorithms for data mining. Most of the previous studies on data mining applications in various fields use the variety of data types range from text to images and stores in variety of databases and data structures. The different methods of data mining are used to extract the patterns and thus the knowledge from this variety databases. Selection of data and methods for data mining is an important task in this process and needs the knowledge of the domain.

REFERENCES


Authors

Architha.S: Presently Architha is pursuing B.E. in Computer Science Engineering at City Engineering College Bangalore, India

A.Kishore Kumar: Presently Kishore is pursuing B.E. in Computer Science Engineering at City Engineering College Bangalore, India

Corresponding Address-

A.Kishore Kumar- #1180, Renuka Nilaya, 2nd main, 7th block, hosakerehalli, bsk 3rd stage, Bangalore- 85
Mobile number- +919845745456
Email id- kishore.kumar.kishore1997@gmail.com

Architha.S- #330/10 ground floor, 23rd cross, 6th block, jayanagar, Bangalore- 70
Mobile number- +918123870278
Email id- architha96@gmail.com