Towards OLAP - Based Data Mining using Multidimensional Database and Fuzzy Decision Trees

Tiruveedula Gopi Krishna, Mohamed Abdeldaiem Abdelhadi, Sabahaldin A.Hussain

Abstract-- In this paper, a new approach for data mining is described, coupling fuzzy multi-dimensional databases and fuzzy data mining systems, and achieving knowledge discovery from imperfect data. An architecture based on fuzzy multidimensional databases is given. It uses these kinds of data repositories to extract relevant knowledge from large data sets from the real world. According to several works that highlighted the great interest of the OLAP framework in the knowledge discovery process, this approach enhances the existing solutions. It provides a way to deal with data from the real world, and to apply flexible operations on data sets stored as multidimensional arrays, generating more understandable fuzzy rules. In recent works an extension of multidimensional database has been defined in order to handle imperfect information and flexible multidimensional queries.

Index Terms-Olap, Fuzzy Rules, Multidimensional Database, Data Mining

I. INTRODUCTION

The purpose of data mining is to discover unknown new information and due to this fact the results are truly useful. The knowledge discovered through data mining must be valid. Appling the data mining techniques on large amounts of varied data could lead also to false information therefore is essential to check the data validity. We could also refer to the datamining process as a step in discovering the information through a set of algorithms and patterns meaningful in the data structures and showing market trends.Data mining discovers patterns within data, using predictive techniques. These patterns play a very important role in the decision making because they emphasize areas where business processes require improvement. Using the data mining solutions, organizations can increase their profitability, can detect fraud, or may enhance the risk management activities. The models discovered by using data mining solutions are helping organizations to make better decisions in a shorter amount of time. Numerical descriptions, leading to more understandable knowledge.

Data mining methods derived from statistical calculation, database administration and artificial intelligence, SQL queries, analysis in multidimensional databases using OLAP systems. They don't replace the traditional methods of statistics, but are considered to be extensions of graphic and statistical techniques.

Manuscript received November, 2013.

Tiruveedula Gopi Krishna, Lecturer, Faculty of Arts and Science, Department of Computer Science, Hoon,, Al-Jufra,Sirte University, Libya. Research Scholor, India.

Dr.Mohamed Abdeldaiem Abdelhadi,Lecturer, Faculty of Arts and Science, Department of Computer Science, Hoon,, Al-Jufra,Sirte University, Libya. Research Scholor. India.

Sabahaldin A.Hussain, Lecturer, Faculty of Arts and Science, Department of Computer Science, Hoon,Al-Jufra, Sirte University, Libya. Research scholar.India.

Typical data structure suitable for data mining contains the observations placed on lines and the variables placed on columns. Domains or range values for each variable must be precisely defined, avoiding as much as possible vague expressions. Line and column format, similar to the spreadsheet file is required for data mining.

Data mining software is separated into two groups:

1. Data mining tools – are providing techniques that can be applied to any business problems.
2. Data mining applications – incorporate techniques inside an application specially built to address business problems. Our life isinfluenced by data mining applications. For example, almost any financial transaction is processed by a data mining application to detect fraud. Increasingly more organizations are using both data mining tools and applications to develop predictive analysis. In order to apply data mining tools, data must go through the following processes:

A. Preprocessing – removing unnecessary data

- Consistency checking (measurement units);
- Detecting and removing erroneous information;
- Removing extreme values (outliners).

B. Data integration – combining variables

C. Transforming variables – through standardization, or by passing to logarithmic scale

D. Separating the database into three categories of data:

- Training category
- Validating category classical one, and provides a framework to handle all classical crisp cases, since fuzzy theory provides means to handle imperfect and classical data.

II. DATA WAREHOUSE AND OLAP TECHNOLOGY:

Data warehouse generalize and consolidate data in multidimensional space. The construction of data warehouse involves data cleaning, data integration and data transformation and can be viewed as an important preprocessing step for the data mining. Moreover, data warehouses provide on-line analytical processing (OLAP) tools for the interactive analysis of multidimensional data of varied granularities, which facilitates effective data generalization and data mining. Many other data mining functions, such as association, classification, prediction and clustering can be integrated with OLAP operations to enhance interactive mining of knowledge at multiple levels of abstraction.

Hence, the data warehouses has become an increasingly important platform for data analysis and on-line analytical processing and will provide and effective platform for data mining. Therefore data warehousing and OLAP form and essential step in the knowledge discovery process.

III. A MULTIDIMENSIONAL DATA MODEL

Data warehouses and OLAP tools are based on a multidimensional data model. This model views data in the form of data cube. In this a data cube allows data to be modeled and viewed in multiple dimensions. It is defined by dimensions and facts.

IV. OLAP OPERATIONS IN THE MULTIDIMENSIONAL DATA MODEL

In multidimensional model, data are organized into multiple dimensions, and each dimension contains multiple levels of abstraction defined by concept hierarchies. This organization provides users with the flexibility to view data from different perspectives. A number of OLAP data cube operations exist to materialize these different views, allowing interactive querying and analysis of the data at hand. Hence, OLAP provides a user-friendly environment for interactive data analysis.

V. OLAP SYSTEMS VERSUS STATISTICAL DATABASES.

Many of the characteristics of OLAP systems, such as the use of a multidimensional data model and concept hierarchies, the association of measures with dimensions, and the notions of roll-up and drill-down, also exist in earlier work on statistical database(SDBs). A statistical database is a database system that is designed to support statistical applications.

VI.FUZZY MULTIDIMENSIONAL DATABASES

Many models have been proposed for multidimensional databases. Roughly speaking, a multidimensional database is set of hypercube (hereafter cubes), defined on dimensions, which may be organized hierarchically. One dimension of particular interest is chosen as the measure whose values are stored in cube cells. Operations are defined to visualize and manipulate cubes (e.g. rotation, selection by slice noddice, roll-up drilldown).

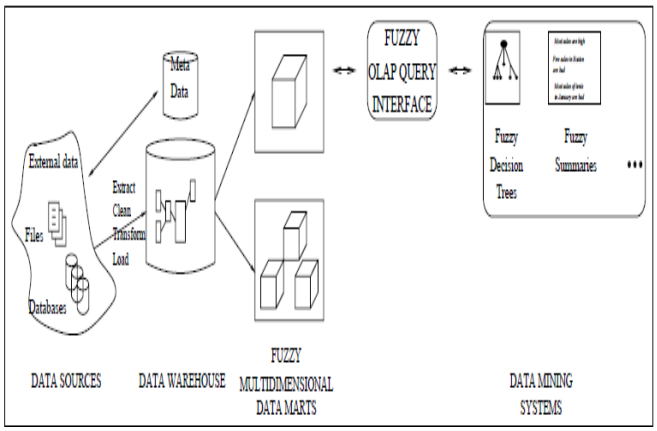

Fig. 1 Architecture

VII. ON-LINE ANALYTICAL DATA MINING SYSTEMS (OLAM)

Data mining and OLAP system are tools for business intelligence. OLAP queries retrieve the database information, at certain levels. OLAP analysis is a deductive process. Based on this hypothesis, data mining is different from OLAP system because it is using its data to discover new patterns. This tool examines the data and interactions between them.

VIII. FUZZY DECISION TREES

The output of the system is a fuzzy decision tree that can be considered as a set of discovered classification rules.Constructing fuzzy decision trees from fuzzy multidimensional databases con-sits in a sequence of OLAP operations of slice and dice and aggregation that we defined above. The algorithm automatically builds fuzzy partitions. This leads to the treatment of numerical data at a higher level of abstraction, and reduces the size of the tree. Themultidimensional database management system may either send only basic in-formation on data, or on the opposite compute complex fuzzy operations and aggregations, or any of the intermediate solutions. In the current version of our system, an interface is implemented that consists in exchanging for each node of the tree the statistics on the data associated with the current node. These data may either be singleton values or intervals, which enables us to construct fuzzy decision trees. These statistics are computed using OLAP queries to extract a sub-cube and fetching statistics on it with aggregation functions. In this framework, multidimensional databases are very interesting since they provide efficient tools for aggregation, especially for the computation of the en-trophy at each step of the construction for the selection of the best attribute to partition data regarding to the class. Moreover, they enable navigation through levels of granularity, thus decision trees can be induced at several levels of gran-ularity. Data mining technology is focused on assessing the predictive power of patterns, this being possible by testing conclusions on a different set of data and by calculating the predictive accuracy. So, we can appreciate that data mining and OLAP are two technologies that complement each other. Data mining could helpanalyze and design the data warehouse by focusing attention on important variables, identifying exceptions and finding interactions between variables. Due to the interconnection between the two technologies, the OLAM systems haveemerged. OLAM systems are also calledOLAP systems for data mining. This type of system integrates OLAP multidimensional processing with extracting knowledge form data, in data mining.

IX. CONCLUSIONS

Data mining tools and applications are helpful in business management, business intelligence, selective marketing, and decision analysis. Data mining is a technology that uses complex and elaborate algorithms in order to analyze and reveal interesting information useful in the analysis made by decision makers. OLAP organizes data into a pattern suitable for the analysts to operate while data mining carries out data analyses and provides the results to the decision makers.

Thus, OLAP enables a model oriented analysis while data mining makes the oriented data analysis easier. In this paper, a new approach for data mining is described, doubling fuzzy multidimensional databases and fuzzy data mining systems and achieving knowledgediscovery from imperfect data. It uses these kinds of data repositories to extract relevantknowledge from large data sets from the real world. Many perspectives are associated with this work. However, our further workconcerns mainly:

1. The enhancement of the current implementation of the system,
2. The integration of other data mining systems (e.g., clustering, prediction),
3. The comparison of differentapproaches managing the complexity of the queries from the data mining system into the multidimensional database management system,
4. The use of discovered knowledge as metadata in order to automatically build interesting cubes from the data warehouse,
5. The use of user expectations to deal with granularity levels in the discovery process and to enhance the system performances

REFERENCES

1. R.Agrawal, T.Imielinski, and A.Swami, Mining association rules between sets of items in large databases. In proc.of AGM SIGMOD conf.,
2. E.F Codd, S.B.Codd, and C.T.salley. providing OLAP to user – analysis: an IT mandate. White thesis, 1993.
3. B.Bouchon-Meunier,M.Rifqi, and S.Bothorel. Fuzzy sets of Systems, 85: 143-153,1996.1. Lungu, Ion and Bâra, Adela. Sisteme informatice executive. Bucharest : Editura ASE, 2007.
4. Velicanu, Manole, et al., et al. Sisteme de baze de date evoluate. Bucharest:Editors ASE, 2009.
5. Chmelar, Petr and Stryka, Lukas. Iterative, Interactive and Intuitive Analytical Data Mining. [Online] 2009.http://www.fit.vutbr.cz/~chmelarp/public/07znalosti%20i3mining.pdf.
6. Gorunescu, Florin. Data Mining Concepts, Models and Techniques: Springer, 2011.
7. Ullman, Jeffrey D. Data Mining Lecture Notes. 2009.
8. Han, Jiawei and Kamber, Micheline. Data mining: concepts and techniques. : Morgan Kaufmann Publishers, 2006.

AUTHORS PROFILE

Tiruveedula Gopi Krishna, Lecturer, Faculty of Arts and Science, Department of Computer Science, Hoon,, Al-Jufra,Sirte University, Libya. Research Scholor. Email: gktiruveedula@gmail.com Research Area: Data Mining published 4 International Journals and 3 National Journals.

Dr. Mohamed Abdeldaiem Abdelhadi, Lecturer, Faculty of Arts and Science, Department of Computer Science, Hoon,, Al-Jufra,Sirte University, Libya. Research Scholor.

Sabahaldin A. Hussain, Lecturer, Faculty of Arts and Science, Department of Computer Science, Hoon,Al-Jufra, Sirte University, Libya. Research scholar.