American Journal of Engineering Research (AJER) 2016
American Journal of Engineering Research (AJER)
e-ISSN: 2320-0847 p-ISSN : 2320-0936
Volume-5, Issue-2, pp-237-240
www.ajer.org
Research Paper Open Access

OLAP Mining Rules: Association of OLAP with Data Mining

Naseema Shaik[1], Dr. Wali Ullah[2], Dr. G. Pradeepni[3]
[1](Research Scholar, Department of Computer Science, KL University, INDIA)
[2] (Assistant Processor, Department of Computer Science, Jazan University, KSA)
[3](Associate Professor, Department of Computer Science, KL University, INDIA)

Abstract: *Data mining and OLAP are powerful decision support tools. The usage of both systems will differ. OLAP systems concentrate on the efficiency of building OLAP cubes. Here No specific algorithms are going to be implement. On the other hand, statistical analyses are traditionally developed for two-way relational databases, and have not been generalized to the multi-dimensional OLAP data structure. Combining both OLAP and data mining may provide excellent solutions. In large data warehouse environments, many different types of analysis can occur. There may also apply more advanced analytical operations on data. Two major types of such analysis are OLAP (On-Line Analytic Processing) and data mining. Rather than having a separate OLAP or data mining engine, OLAP mining rules has integrated OLAP and data mining capabilities directly into the database server. This paper provides a brief introduction to these technologies, and more detail can be found in these products' respective documentation.*

Keywords: *Data mining, OLAP systems, OLAP cubes, Association rules, OLAP mining.*

I. INTRODUCTION

The OLAP and data mining (OLAP mining) is a mechanism which integrates on-line analytical processing (OLAP) with data mining so that mining can be performed in different portions of databases or data warehouses and at different levels of abstraction at user's finger tips. With rapid developments of data warehouse and OLAP technologies in database industry, it is promising to develop OLAP mining mechanisms. With our years of research into data mining, an OLAP-based data mining system, DBMiner, has been developed, where OLAP mining is not only for data characterization but also for other data mining functions, including association, classification, prediction, clustering, and sequencing. Such integration increases the flexibility of mining and helps users find desired knowledge. In this paper, we introduce the concept of OLAP mining and discuss how OLAP mining should be implemented in a data mining system by using association rules.

II. OLAP AND DATA MINING ARE USED TO SOLVE DIFFERENT KINDS OF ANALYTIC PROBLEMS

OLAP summarizes data and makes forecasts. For example, OLAP answers questions like "What are the average sales of policies, by area and by year?" where as Data mining discovers hidden patterns in data. Data mining operates at a detail level instead of a summary level. Data mining answers questions like "Who is likely to buy policies in the next six months, and what are the characteristics of these likely buyers?" OLAP and data mining can complement each other. For example, OLAP might pinpoint problems with sales of policies in a certain area. Data mining could then be used to gain insight about the behavior of individual customers in the region. Finally, after data mining predicts something like a 5% increase in sales, OLAP can be used to track the net income.

III. OLAP MINING OVERVIEW & ARCHITECTURE

"Online Analytical Processing Provides you with a very good view of what is happening, but cannot predict what will happen in the future or why it is happening where as Data Mining is a combination of discovering techniques and prediction techniques [1]."

The architecture which proposes in this paper fulfills several important, and often interrelated, goals:

1. **Modularity**: All modules, which adhere to a predefined interface, can interact seamlessly Interoperability. The system has to work with a wide array of databases and storage models. The integration of multiple database systems based on wrapper modules needs to be supported.
2. **Scalability:** OLAP requires consistent reporting performance, independent of the size of the underlying database or its dimensionality.
3. **Extendibility:** We want to add additional modules without rebuilding the system [2].

The data mining and OLAP are powerful tools to support decision making. However, people use them separately for years: OLAP systems focus on efficiency to build cubes OLAP, and applied to any algorithms for mining statistical data, on the other hand, developed a traditional statistical analysis for two-way relational databases, and have not been circulated to the structure of the multi-dimensional data OLAP. May combine data mining and OLAP to provide excellent solutions. OLAP is completely different from its predecessor, for an online transaction processing (OLTP) systems. OLTP focuses on the automation of data collection procedure. Keeping detailed data, consistent and modern, is the most important condition for the application of OLTP.

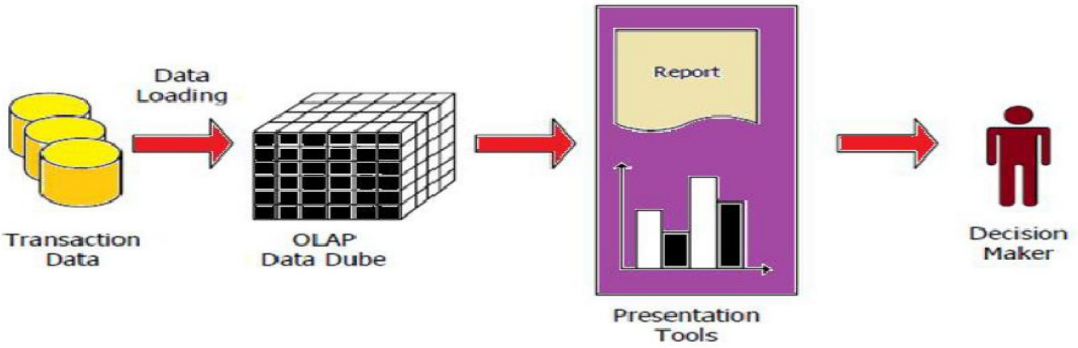

Fig1: OLAP Architecture

Component-based OLAP systems offer a number of benefits both to the user and the developer. The user can choose from different suppliers and combine different query optimization strategies and query evaluation algorithms [3]. Extensible OLAP systems offer a major benefit to developers, as well. We will show how to decompose an OLAP system into functional units, which can communicate using a "software bus". We propose a data model based on sets and vectors for the communication.

A number of approaches exist to the storage of data in data warehouses. [4] Contrasts the performance of value based (ROLAP) and multidimensional (MOLAP) implementations. The data-structures used include grid-files [5], B*-trees, R*-trees [1], X-trees, HB-trees [6], Gist [7], arrays and sets-based data-structures.

OLAP systems have a structured architecture based on the following essential components:

1. Database - the data source used for OLAP analysis. As database can use a relational database to ensure our multidimensional storage facilities, a multidimensional database, a data warehouse, etc.

2. OLAP server - the one that manages multidimensional data structure and at the same time a link between the database and OLAP customer.

3. OLAP customer - are those that provide data mining applications but also supports the generation of results

(Graphs, reports, etc.). There are several options in OLAP data could be stored and processed. Thus, depending on the method of organizing and storing data, there may be three options [2]:

4. Client Files - data is stored locally on a client computer as files are organized, on which operations can be applied to analyze the processing and transformation. This organization of data has some drawbacks of which we can enumerate: the amount that can be processed is indulged reduced time to processing information is quite high, the data shows a poor security, lack of advanced multidimensional analysis.

5. Relational databases - this arrangement is used when the data comes as a relational DBMS and data warehouse is a repository be implemented virtually or using a relational model.

6. Databases multidimensional - in this case, the data are organized into a data warehouse on a dedicated server, which is called multidimensional server.

The figure below shows the architecture of OLAP systems, which vary depending on how data storage and processing of their type, but generally on how one can identify three levels of data: the data sources, OLAP server and the presentation of data or interface user.

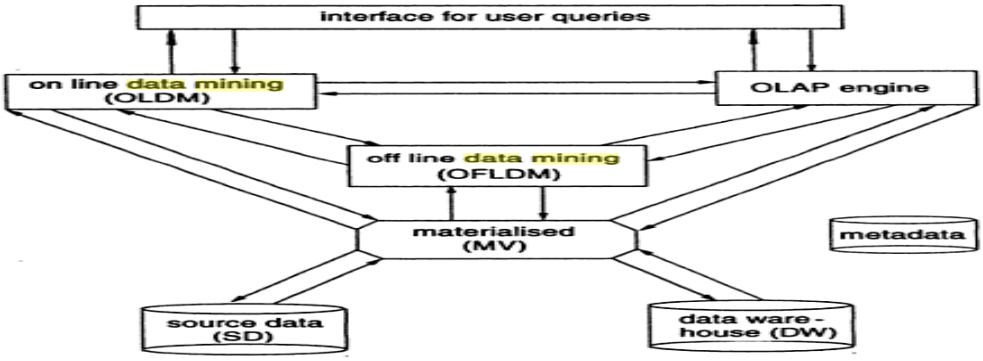

Fig 2: An Integrated view of OLAP Mining Architecture

Some researchers began to generalize some data mining concepts on OLAP cubes in recent years. These works include the cube grade problem the constrained gradient analysis, and data-driven OLAP cube exploration [8]. We will review these studies briefly in this section.

It is a generalized version of association rule. Two important concepts in association rule are support and confidence. Let us take the market basket example. Support is the fraction of transactions that contains a certain item (bread and butter), and confidence is that the proportion of transactions that contains another item B given that these transactions contain A which declare that the association rule can also be viewed as the change of the count aggregates when imposing another constraint, or in OLAP terminology, making a drill-down operation on an existing cube cell. They think other aggregates like sum, average, max, and min can be studied in addition to the count. Also, other OLAP operations, like roll-up and one-dimension mutation can be incorporated [9]. They argued that the cube grade could support the "what if" analysis better, and they introduced two query languages to retrieve the cube grade sets.

IV. IMPLEMENTATION OF OLAP MINING BY ASSOCIATION

Association rule mining finds interesting relationships in data. The goal of associative rule data mining is to find all associative rules that have high confidence (Strong Rules) in the data set. In meteorological application, association mining used to find the relationship between the weather elements and natural events, weather and disaster prediction [8], and multi-station atmospheric data analysis [10].

Table below illustrates some useful rules extracted from New York weather data ordered by confidence.

Associations rules for New York City weather data

#	Rule	Conf.
1	[RH=mid Temp=warm Wind=Moderate] ==> [Rain=no rain]	0.99
2	[RH=high Temp=warm] ==> [Rain=no rain]	0.99
3	[Temp=warm Wind=Moderate] ==> [Rain=no rain]	0.99
4	[month = 2]==> [temp = cold]	0.96
5	[month = 1] ==> [temp = cold]	0.96
6	[month = 12] ==> [temp = cold]	0.95
7	[Wind=Light] ==> [Rain=no rain]	0.91
8	[Wind=light Temp=cold rain]==>[RH=moderate]	0.91
9	[Rain= Heavy Rain]==>[Temp = cold]	0.88
10	[Temp = cold]==>[Wind = Moderate]	0.74
11	[RH= low Wind = Moderate Temp=warm] ==> [Rain=Light Rain]	0.65
12	[Wind = Moderate] --> [RH = mid]	0.60

Rules #1, #2, #3, #7 and #11, can be used to predict rainfall. For example from rule #1 we understand that there is no rain tomorrow if today is warm (temperature between 16 °C and 23 °C), wind speed is moderate (13-30 km/h) and relative humidity is mid (between 56.5 - 76.0). Also Rule #11 could be used for rain prediction, it means that if the relative humidity today is low (below 36), wind speed is moderate and temperature is warm then, rain tomorrow maybe light (< 2.5 millimeters per hour). Rules #4, #5 and #6 provide with better understanding for Gaza city weather. These rules give us an indication that cold season includes December, January and February.

V. CONCLUSION

OLAP Mining does not provide any formal or standard technique to be modeled. Each vendor defines their own approach regarding the needs of respective end users. However, there is a general model used on Data Warehouses called Star schema but it cannot model all the appropriate conceptual issues and problems like information loss is very often. A similar approach, which overpasses the information loss problem of the Star schema, is the Snowflake schema also a common database model for Data Warehousing but as most of computer scientists claim it is a logical view rather a conceptual view of the database model.

REFERENCES

[1] Cop pock D. S. "Data Mining and Modeling: So you have a Model, Now What?" DM Review Magazine, Feb 03.
[2] Salton, G. and McGill, M. Introduction to Modern Information Retrieval, McGraw-Hill Book Company, New York 1983.
[3] Lungu I., Bâra A., Sisteme informatice executive, Editura Ase, Bucureşti, 2007.
[4] Central Intelligence Agency, "CIA World Fact book 2011," Mobile Reference. 2010.
[5] Li X., Plale N., Vijayakumar R., Ramachandran S., Graves H., "Conover. Real-time storm detection and weather forecast activation through data mining and events processing, "To appear Earth Science Informatics, H.A. Babaie, Ed., Springer. 2008.
[6] Berkhin P., "Survey of clustering data mining techniques, Accrue Software, San Jose", CA, Tech. Rep., 2002.
[7] Dong-Jun S., Breidenbach, J. P., "Real-Time Correction of Spatially Nonuniform Bias in Radar Rainfall Data Using Rain Gauge Measurements," Journal of Hydrometeorology, vol. 3, no. 2, pp. 93- 111. 2002.
[8] R. Kimball. The Data Warehouse Toolkit: Practical Techniques for Building.Dimensional Datawarehouses. John Wiley & Sons, Inc., 1996.
[9] J. Widom. Research problems in data warehousing. In Conference on Information and Knowledge Management, 1995.
[10] Nandagopal S., Karthik S.Arunachalam V."Mining of Meteorological Data Using Modified Apriori Algorithm," European Journal of Scientific Research, No.2, pp.295-308. 2010.