Research Paper

Volume : 2 | Issue : 3 | Mar 2013 • ISSN No 2277 - 8179

Computer Science

KEYWORDS : Healthcare, Clinical decision support system, OLAP, Data mining

Integrated approach of OLAP and Data Mining for Decision support

Ms. Alpa R. Patel — Research Scholar of JJT University, Jhunjhunu, Rajasthan-333001

Dr. Jayesh M. Patel — Associate Professor, MCA Department, Acharya Motibhai Institute of Computer Studies, Ganpat University, Kherva, Gujarat

ABSTRACT

The healthcare industry collects huge amounts of data which are not turned into useful information for effective decision making. Decision support systems (DSS) can now use advanced technologies such as On-Line Analytical Processing (OLAP) and data mining to deliver advanced capabilities. This paper presents a model for clinical decision support system which combines the strengths of both OLAP and data mining.

1. Introduction

The healthcare industry is under pressure to lower cost and improve service quality. Oftentimes, information produced is excessive, disjointed, incomplete, inaccurate, in the wrong place, or difficult to make sense. A critical problem facing the industry is the lack of relevant and timely information. As information costs money, it must adopt innovative approaches to attain operational efficiently. Decision Support Systems (DSS) have been developed to overcome these limitations. This paper presents a model for clinical decision support system based on OLAP and data mining. It provides a rich knowledge environment which is not achievable by using OLAP or data mining alone.

2. The Model

OLAP uses several preprocessing operations such as data cleaning, data transformation, data integration, its output can serve as valuable data for data mining. OLAP operations (e.g., drilling, dicing, slicing, pivoting, filtering) enable users to navigate data flexibly, define relevant data sets, analyze data at different granularities and visualize results in different structures. Applying these operations can make data mining more exploratory. The motivation for an integrated model, OLAP with data mining, is the concept hierarchy. Data in OLAP and decision tree are organized into multiple dimensions where each dimension contains multiple levels of abstraction defined by the concept hierarchy. The concept hierarchy is illustrated in Fig. 1, where each member has one root and all members between roots have parents and every branch ends with a leaf member.

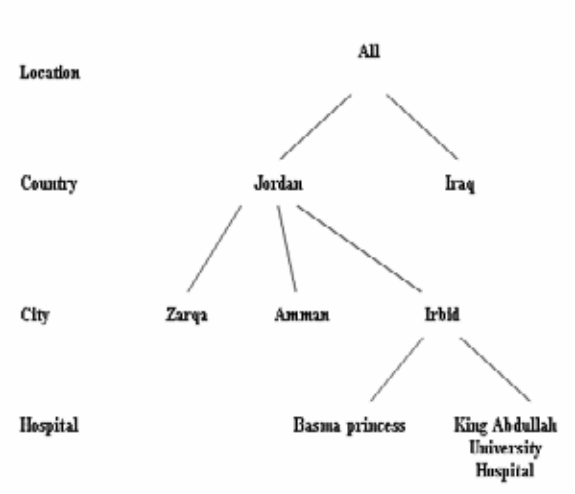


Fig.1 Concept hierarchy for the dimension location

3. Research Questions

This research demonstrates: "How can integrated OLAP with data warehousing, to provides advanced decision support System compared to using OLAP or data warehousing alone"?

The research questions listed in Table 1 are used for this purpose. They can be answered by using the integrated model. Fig. 2 shows the architecture of the integrated model (OLAP with data mining) comprising of several components. The system is divided into two parts: Serverside – for building the integrated model, and Client-side – for accessing queries and presenting results (Fig. 2). It uses OLAP operations and the decision tree mining algorithm C4.5. The test data validates the effectiveness of the model.

R1. How does OLAP with data mining enhance real time indicators like bottlenecks?
R2. How does OLAP with data mining provide improved visualization to uncover patterns/trends that are likely to be missed?
R3. How does OLAP with data mining uncover more subtle patterns in data over capabilities provided by OLAP or data mining alone?

Table.1 Research Questions

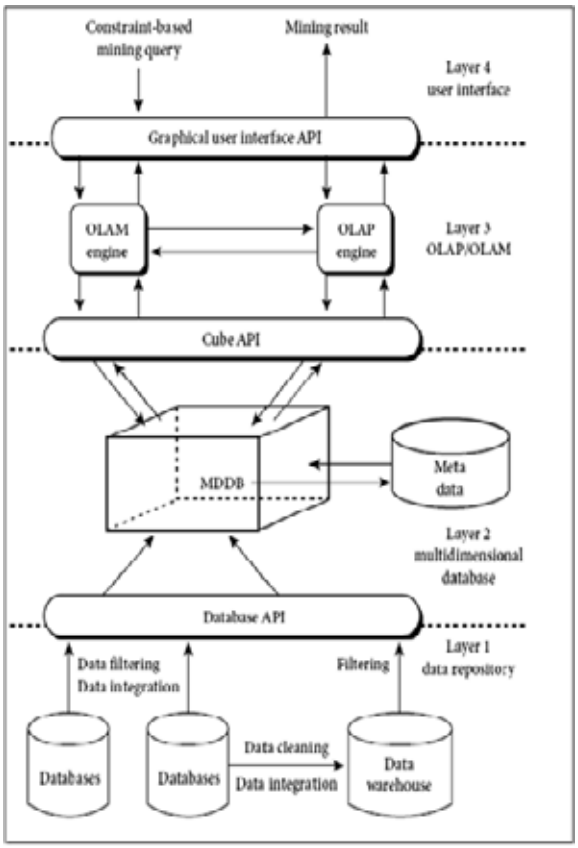


Fig.2 Integration of OLAP with data mining architecture

4. System Design

A data cube is first created then the data mining process is start-

ed. The cube preserves the information and allows browsing at different conceptual levels. It serves as the data source for the data mining task. Data mining can be performed on any level or dimension of the cube. After the model is built it is stored in the OLAP cube. Each dimension represents the rule corresponding to a node in the decision tree mining model (Fig.3). OLAP operations explain the different states of the system. The data for this study is taken from UCI Repository of Machine Learning Databases. The data comprise of Pima Indian Diabetes database, Post Operative Recovery database and Liver Disorder database. As the data is declassified, we have added several dummy attributes such as patient, doctor and hospital information.

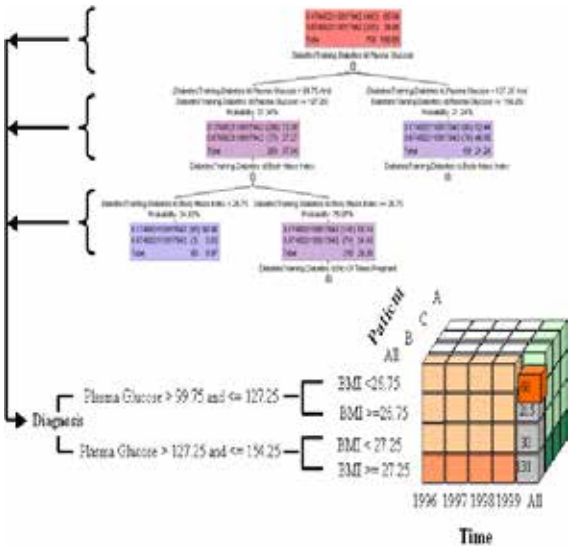


Fig.3: A logical view of representation of decision tree mining model in an OLAP cube

A two-year sample dataset (1997-1998) is created to mine for knowledge discovery. Information on entities and their attributes and relationships are fed into the data warehouse. Its design is based on the star schema (Fig. 4).

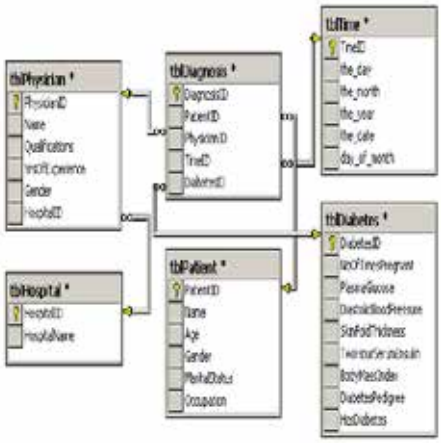


Fig.4 Data warehouse star schema design

5. System Implementation

The system can predict the future state and generate useful information for effective decision-making. It can answer all the research questions listed in Table 1.

R1. The integrated model enhances real time indicators by using information on hospital room utilization for postoperative recovery patients. It allows hospital administrators to discover any bottlenecks that might exist. It allows them to solve problems related to hospital room utilization. The results show that a total of 6 patients are likely to be warded in Hospital A. The administrator can use this information to allocate rooms based on their characteristics. For patients over 60 years, a decision may be made to ward them in senior citizen's ward or transfer them to another hospital. This indicator is useful for performing "what-if" analysis on hospital room availability.

R2. The integrated model improves information visualization. It discovers overall trends that are likely to be missed by using OLAP or data mining alone.

R3. With data mining, doctors can predict patients who might be diagnosed with diabetes. OLAP provides a focused answer using historical data. However, by combining, we can optimize existing processes and uncover more subtle patterns, for example, by analyzing patients' demographics. that are likely to be warded at Hospital A. Slicing operation on a data cube to identify post-operative recovery patients who are likely to be warded. The system allows users to perform advanced data analysis and ad hoc queries. It also generates reports in multiple formats.

6. Conclusion

This paper has presented a DSS based on OLAP with data mining. The system is powerful because (1) it discovers hidden patterns in the data, (2) it enhances real time indicators and discovers bottlenecks and (3) it improves information visualization.

REFERENCE

[1] Blake, C.L. & Merz, C.J., UCI Repository of Machine Learning Databases, University of California, Department of Information and Computer Science, 1998. Available: http://www.ics.uci.edu/~mlearn/MLRepository.html | [2] Donald, J.B., John, W.F., Alan, R.H., James, S., Healthcare Data Warehousing and Quality Assurance, IEEE Computer, pp.56-65, 2001, December. | [3] Fayyad, U., Gregory, P.-S. and Smyth, P., From Data Mining to Knowledge Discovery in Databases, AI Magazine, 37(3), pp.37-54, 1996. | [4] Fong, A.C.M, Hui, S.C., and Jha, G., Data Mining for Decision Support, IEEE IT Professional, 4(2), pp.9-17, March/April, 2002. | [5] Han, J., OLAP Mining: An Integration of OLAP with Data Mining, Proceedings of 1997 IFIP Conference on Data Semantics (DS-7), Leysin, Switzerland, pp.1-11, 1997, October. | [6] Parseye, K., OLAP and Data Mining: Bridging the Gap: Database Programming and Design, 10, pp.30-37, 1998. | [7] Ralph, K. and Margy, R., The Data Warehouse Toolkit. The Complete Guide to Dimensional Modeling (2nd ed.), Canada: John Wiley & Sons, Inc, 2002. | [8] Robert, S.C., Joseph, A.V. and David, B., Microsoft Data Warehousing: Building Distributed Decision Support Systems, London: Idea Group Publishing, 1999. | [9] Surajit, C. and Umeshwar, D., An Overview of Data Warehousing and OLAP Technology, ACM Sigmod Record, 26(1), pp.65-74, 1997. |