

An Study on Industrial Data Warehouse and Data Mining Technique

Ankit Yadu¹, Rahul Kumar Chawda²

¹Student of MCA, ²Assistant Professor

Department of Computer Science, Kalinga University, Raipur

Abstract - Increasing use of computers in every business process, leads to accumulation of data of an organization, demanding the need of sophisticated data handling techniques. Many data handling concepts have evolved that support data analysis, and knowledge discovery from the historical data. Data warehouse and Data mining technology is playing an important role in the area of data analysis for knowledge discovery. Various data mining techniques are available to mine the data from data warehouse. These techniques typically address the four basic applications such as data classification, data clustering, association between data and finding sequential patterns between the data.

Keyword - Data Warehouse , Data Mining , Software

I. INTRODUCTION

Use of computers is practically present and increasing in every business process, leading to accumulation of valuable data giving rise to voluminous historical data of an organization. These historical data are the reflection of business status and performance of an organization. This phenomenon is demanding the need of sophisticated data handling tools at all levels of business organization. Many data handling concepts have evolved that support data analysis, and knowledge discovery from the historical data. Various tools and applications are available for data processing, transaction processing, information processing and now for knowledge discovery. As we move from data processing to knowledge discovery the nature of data and type of data handling methods differ as stated below.

A. Transactional processing

- i). high user intervention
- ii). current data
- iii). small data volume
- iv). Structured queries are designed as the nature of output is known

B. Information processing

- i). moderate user intervention
- ii). current data
- iii). periodically frequent data
- iv). data for one business year or one project

C. Data Warehouse and Data Mining

- i). very little user intervention
- ii). historical data
- iii). data can be for the number year as that of the age of an organization

- iv). Data handling methods are highly automated.

The research study is focused in the area Data warehouse and Data Mining, addressing two major issues of data classification using decision tree, 1) classification of variables of unknown or uncertain characteristics, 2) creating classification by combining multiple variables. Data warehouse is a database created by combining data from multiple databases for the purposes of analysis . Data Mining is the analysis of (often large) observational data sets to find unsuspected relationships and to summarize the data in novel ways that both understandable and useful to the data owner . These techniques typically address the four basic applications such as data classification, data clustering, Association between data and finding sequential patterns between the data. Various data mining techniques are available to mine data from data warehouse. Data mining tools are automated software that allows users to perform detailed mathematical and statistical calculations on detailed data warehouse data to detect trends, identify patterns and analyze data. These data thus have proved the basis of accurate decision making in the fields like retail, banks, fraud detection, customer analysis etc.

II. LITERATURE REVIEW

A Data Warehouse (DW) on the other end is a database that is designed for facilitating querying and analysis. A separate computer databases began to be built that was specifically designed to support management information and analysis purposes. These data warehouses were able to bring in data from a range of different data sources, such as mainframe computers, minicomputers, as well as personal computers and office automation software such as spreadsheets and integrate this information in a single place. This capability, coupled with user friendly reporting tools, and freedom from operational impacts has led to a growth of this type of computer system.

Data Warehousing Architecture and its Components, the popular 3-tier architecture consists of several components. The 3-tier architecture comprises of warehouse server at tier 1, OLAP-engine for analytical processing at tier 2, and a client containing reporting tools, visualization tools, data mining tools and querying tools at tier 3.

Figure 1 shows a typical data warehousing architecture.

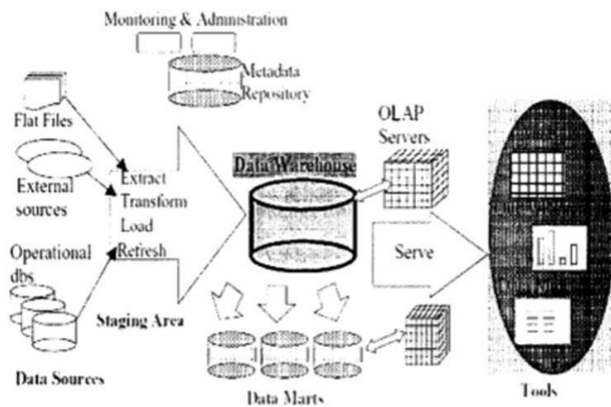


Figure 1: A typical data warehousing architecture Metadata

Metadata - Is to the data warehouse what the card catalogue is to the traditional library. It serves to identify the content and location of data in the warehouse. Metadata is a bridge between the data warehouse and the decision support application. Meta data may also contain data extraction/transformation history, column aliases, data warehouse table sizes, data communication/modeling algorithms and data usage statistics.

OLAP Engine - Operational database like retail database, customer database, or the reservation system of an airline are most often used to answer well-defined and repetitive queries such as "What is the total price of the products in this basket", "What is the address of customer Smith"? Such databases have to support a large number of transactions consisting of simple queries and updates on the contents of the data. This type of database usage is called online transaction processing (OLTP).

III. CONCLUSION

Various algorithms that address to classification on large data sets have proved to be efficient in classifying the variables of known or certain characteristics. Tools developed using these algorithm are more automated with less or without the intervention of an experienced user. Due to which the results or meanings produces out of these tools carry less realistic meaning when it comes to analysis of variables of unknown or uncertain characteristics. These algorithms do not address any clear method to create a class by combining multiple correlated variables in the real world. Introduction

Following issue are arrived during review on various paper

- 1) Do not have a clear method to combine multiple variables which has greater decision meaning in real world.
- 2) Pruning techniques used sometimes may ignore those variables which may have more influence in real world.
- 3) Very less scope for user intervention

IV. REFERENCES

- [1]. Ahmad Irtishad and Azhar Alman, "Data Warehousing in Construction, Advancing Engineering, Management and

Technology", Third International Conference in the 21s' Century (C1TC-III), Athens, 2005.

- [2]. Ahmad Irtishad, "Data Warehousing in Construction Organizations", Construction Congress VI, American Society of Civil Engineers, 2000, pp. 194-203.
- [3]. Goebel Michael, Gruenwald Le, "A survey of data mining and knowledge discovery software tools", ACM SIGKDD Explorations Newsletter, Volume 1, Issue 1, June 1999, p 20 -33
- [4]. Hand David, Mannila Heikki, Smyth Padhraic, Principles of Data Mining, Cambridge, MA, USA, MIT Press, 2004, pp. 141-157, 417-418.
- [5]. IBM, "Advanced Data Warehouse Multidimensional Modeling with DB2 OLAP Server", Eastern Economic Edition, Prentice Hall of India, New Delhi, 2004, P 1.
- [6]. Jen-Rong Lee, Sung-Lin Hsueh, Hung-Ping Tseng, (2008), Utilizing Data Mining to Discover Knowledge in Construction Enterprise Performance records, Journal of Civil Engineering and Management 2008, 14(2): 79-84
- [7]. Jogleker. P.T., "Practical Information for Quantitative Surveyors, Property Valuers, Architects, Engineers and Builders" edition 5, Pune Vidhyarthi Gruh, 2000.
- [8]. Kang Leen-Seok, Paulson C. Boy, Kwak Joong-Min, Kim Chang-Hak, "Business Breakdown Structure for Construction Management and Web-Based Application System", ITcon Vol. 10, 2005, p 169 – 191.
- [9]. Lipman Robert and Kent Reed, "Using VRML In Construction Industry Applications", Web3D - VRML 2000 Symposium, Monterey, CA, February 21-24, 2000, p 119-124.
- [10]. Martin Fischer, Thomas Froese, "Examples and Characteristics of Shared Project Models", Journal of Computing in Civil Engineering, Volume 10, Issue 3, 1996, p 174-182.
- [11]. Moselhi, O. and Hassanein, A., "Data acquisition and analysis for highway construction using geographic information systems", Canadian Journal of Civil Engineering, (30):533-542, 2003.
- [12]. Treiguts Egmonts, "Constructing New Attributes for Algorithms of Decision Trees Induction", Information Technology and Management Science, Scientific Proceedings of Riga Technical University, 2002. 142
- [13]. Van Leeuwen JP, Fridqvist S, "Supporting Collaborative Design by Type Recognition and Knowledge Sharing", ITcon, Vol. 7, Special Issue ICT for Knowledge Management in Construction, 2002, p 167-182, <http://www.itcon.Org/2002/11>
- [14]. Viswanathan P.K., "Glimpses into Application of Chi-Square Tests in Marketing" http://davidmlane.com/hyperstat/viswanathan/chi_square_marketing.html
- [15]. Zhang Jun, Silvescu Adrian, and Honavar Vasant, "Ontology-Driven Induction of Decision Trees at Multiple Levels of Abstraction", SARA 2002, LNAI 2371, 2002, p 316-323.