RESEARCH ARTICLE OPEN ACCESS

An Integrated Architecture for Intelligent Manufacturing Systems in Data Mining

*A.Arokia Marshal Roach ,**G.Raja Raja Cholan
* Assistant Professor, Department of BCA, Immaculate College for Women, Viriyur.
** Assistant Professor, Dept.of CSE, Prist University, Tanjore.

Abstract:

Data Mining is a concept that is taking off in the commercial sector as a means of finding useful information out of gigabytes of data. While products for the commercial environment are starting to become available, tools for a scientific environment are much rarer (or even non-existent). Yet scientists have long had to search through reams of printouts and rooms full of tapes to find the gems that make up scientific discovery. To understand the data warehouse, it is important for you to realize that it is not a single object. It is more of a strategy or a process, an integration of various support systems and programs that are knowledge based. The goal of using a data warehouse is to allow businesses and organizations to make strategic decisions.

I. INTRODUCTION

Users are in an age often referred to as the information age. In this information age, because we believe that information leads to power and success, and thanks to sophisticated technologies such as computers, satellites, etc., users have been collecting tremendous amounts of information. Initially, with the advent of computers and means for mass digital storage, users started collecting and storing all sorts of data, counting on the power of computers to help sort through this amalgam of information. Unfortunately, these massive collections of data stored on disparate structures very rapidly became overwhelming. This initial chaos has led to the creation of structured databases and database management systems (DBMS). The efficient database management systems have been very important assets for management of a large corpus of data and especially for effective and efficient retrieval of particular information from a large collection whenever needed. The proliferation of database management systems has also contributed to recent massive gathering of all sorts of information. Today, users have far more information than we can handle: from business transactions and scientific data, to satellite pictures, text reports and military intelligence. Information retrieval is simply not enough anymore for decision-making. Confronted with huge collections of data, users have now created new needs to help us make better managerial choices. These needs are automatic summarization of data, extraction of the "essence" of information stored, and the discovery of patterns in raw data.

II.EXISTING SYSTEM:

The basic problem here is the translation of the underlying information into the format of the data model used by the warehousing system. A component that translates an information source into a common integrating model is called a *translator* or wrapper. Another problem in data extraction is *update detection*, i.e. monitoring the information source for changes to the data that are relevant to the warehouse, and propagating those changes to the integrator. Depending on the facilities that provide for update detection, the information sources can be classified into several types, so that notification of changes of interest can be programmed to occur automatically. Each of these types of information sources provides particular research problems for update detection. Clearly, the functionality of a wrapper/monitor component depends on the type of the source, and on the data model used by the source. Therefore, developping techniques that automate the process of implementing wrapper/monitor components is an important research problem.

III.PROPOSED SYSTEM:

The task of the integrator is to receive update notifications from the wrapper/monitors and reflect these updates in the data warehouse. Now, the data in the warehouse can be seen as a set of materialised views, and thus the task of the integrator can be seen as materialised view maintenance. However, there are two reasons why conventional techniques cannot be applied directly here, each giving rise to interesting research problems. The first reason is that warehouse views are often more complicated than usual views of relational systems, and may not be expressible using standard languages, such as SQL. For example, typical warehouse views may contain historical data or highly aggregated and summarized information, while the underlying sources may not contain such information. So relevant areas of research here include temporal databases, and efficient view maintenace in the presence of aggregate and summary information.

The second reason why conventional techniques cannot be applied directly here is that the base relations of the warehouse views reside at the information sources. As a result, the system maintaining the warehouse views is only loosely coupled with the systems maintaining the base data. In fact, the underlying information sources simply report changes but do not participate in warehouse views maintenance. In this context, keeping the warehouse views consistent with the base data is a difficult problem, and sophisticated algorithms must be used for view maintenance.

IV.CLUSTERS FROM A DATA CUBE

Clustering groups all the value of one dimension by the values of the second dimension. In DBMiner, only two cube dimensions can be chosen in a mining session since the clustering space is 2-dimensional. The underlying algorithm used in DBMiner is the *k*-means method. For detailed information about the *k*-means method, see the on-line documentation.

Number of clusters refers to *k* in the *k*-mean algorithm. If the number of clusters requested exceeds the number of points in the plane, an error message will be issued and the clustering process is stopped.

Dimension weights refers to the coefficients for each dimension. The default value is 1.00 but can be decreased if you want a particular dimension to be relatively less influential with respect to the other dimension. Because this scaling reflects the relative different in influence, it is only calibrated between 0.01 - 1.00.

Max clustering passes refers to the number of passes in the *k*-mean algorithm.

Filter threshold ensures that cells in a data cube containing no records are not included in the clusters. This threshold can be raised to exclude more cells and thus reduce the number of points on a 2-dimensional plane.

Step(a):

In order to input such a query, follow the menu route Mining -> Mining Wizard to invoke the mining wizard. The mining wizard box will appear as follows.

V. CONCLUSIONS

Data Mining is a new term and formalism for a process that has been undertaken by scientists for generations. The massive increase in the volume of data collected or generated for analysis with the use of computers has made it an essential tool. However, despite the more formal approach, Data Mining is something that scientists perform on an ad hoc basis and can easily adapt to. Many of the methods used for the analysis of the data were originally developed to process scientific data and are used unchanged. As a final point, the biggest of all, the Internet, is becoming more and more important, and while there is useful information, extracting that from the terabytes being added daily is an enormous task. The techniques of Data Mining are applicable here more than any other domain. However, to make use of it takes time, effort and, above all, people with a knowledge of the field, to differentiate the true solutions from the infeasible.

VI. REFERENCES

[1] Chieh-Jen Cheng, Chao-Ching Wang, Wei-Chun Ku, Tien-Fu Chen , and Jinn-Shyan Wang, “Scalable High-Performance Virus Detection Processor Against a Large Pattern Set for Embedded Network Security” Commun. vol. 51, pp. 62–70,2011.

[2] O. Villa, D. P. Scarpazza, and F. Petrini, “Accelerating real-time string searching with multicore processors,” Computer, vol. 41, pp. 42–50,2008.

[3] D. P. Scarpazza, O. Villa, and F. Petrini, “High-speed string searching against large dictionaries on the Cell/B.E. processor,” in Proc. IEEE Int. Symp. Parallel Distrib. Process., 2008, pp. 1–8.

[4] D. P. Scarpazza, O. Villa, and F. Petrini, “Peak-performance DFA based string

matching on the Cell processor," in Proc. IEEE Int. Symp. Parallel Distrib. Process., 2007, pp. 1–8.

[5] L. Tan and T. Sherwood, "A high throughput string matching architecture for intrusion detection and prevention,"in Proc. 32nd Annu. Int. Symp. Comput. Arch., 2005, pp. 112–122.

[6] S. Dharmapurikar, P. Krishnamurthy, and T. S. Sproull, "Deep packet inspection using parallel bloom filters," IEEE Micro, vol. 24, no. 1, pp.52–61, Jan. 2004.

[7] R.-T. Liu, N.-F. Huang, C.-N. Kao, and C.-H. Chen, "A fast string matching algorithm for network processor-based intrusion detection system," ACMTrans. Embed. Comput. Syst., vol. 3, pp. 614–633, 2004.

[8] F. Yu, R. H. Katz, and T. V. Lakshman, "Gigabit rate packet pattern matching using TCAM," in Proc. 12th IEEE Int. Conf. Netw. Protocols, 2004, pp. 174–178.intrusion detection system," ACMTrans. Embed. Comput. Syst., vol. 3, pp. 614–633, 2004.

[9] R. S. Boyer and J. S. Moore, "A fast string searching algorithm,"Commun. ACM, vol. 20, pp. 762–772, 1977.

[10] V. Aho and M. J. Corasick, "Efficient string matching: An aid to bibliographic search," Commun. ACM, vol. 18, pp. 333–340, 1975