

## **DESIGN AND IMPLEMENTATION OF AUTOMATED DATA MINING USING INTELLIGENT AGENTS IN OBJECT ORIENTED DATABASES**

V. Saravanan<sup>1</sup> and Dr. K. Vivekanandan<sup>2</sup>

*1. Dept of Computer Science & Engineering, Bharathiar University, Coimbatore – 641 046, INDIA, [tvssaran@hotmail.com](mailto:tvssaran@hotmail.com)*

*2. School of Management, Bharathiar University, Coimbatore – 641 046, INDIA,*

*[vivek1958@hotmail.com](mailto:vivek1958@hotmail.com)*

**Abstract:** Data Mining is the process of posing queries and extracting useful information, patterns and trends previously unknown from large quantities of data. Agents are defined as software entities that perform some set of tasks on behalf of users with some degree of autonomy. This research work deals about developing a automated data mining system which encompasses the familiar data mining algorithms using intelligent agents in object oriented databases and proposing a framework. Because the data mining system uses the intelligent agents, a new user will be able to interact with the data mining system without much data mining technical knowledge. This system will automatically select the appropriate data mining technique and select the necessary fields needed from the database in a right time without expecting the users to specify the specific technique and the parameters. Also a new framework is proposed for incorporating intelligent agents with automated data mining. One of the major goals in developing this system is to give the control to the computer for learning automatically by using intelligent agents for the exploratory data mining.

**Key words:** Data Mining, Cluster analysis, Intelligent agents

## 1. 1. INTRODUCTION

Data Mining is the process of posing queries and extracting useful information, patterns and trends previously unknown from large quantities of data [Thu, 00]. It is the process where intelligent tools are applied in order to extract data patterns [JM, 01]. This encompasses a number of different technical approaches, such as cluster analysis, learning classification and association rules, and finding dependencies. Agents are defined as software entities that perform some set of tasks on behalf of users with some degree of autonomy. This research work deals about developing a automated data mining system which encompasses the familiar data mining algorithms using intelligent agents in object oriented databases and proposing a framework. Because the data mining system uses the intelligent agents, a new user will be able to interact with the data mining system without much data mining technical knowledge. This system will automatically select the appropriate data mining technique and select the necessary field needed from the database at the appropriate time without expecting the users to specify the specific technique and the parameters. Also a new framework is proposed for incorporating intelligent agents with automated data mining. One of the major goals in developing this system is to give the control to the computer for learning automatically by using *intelligent agents*.

## 2. REVIEW OF LITERATURE

How the data mining systems uses right algorithms and correct data types for the data analysis is the big issues in the data mining [Thu, 00]. Many familiar data mining systems are also analysed in this research project. This includes DBMINER 2.0, Polyanalyst, Miner3D, Knowledge Miner etc.,. Even though all these systems supports for the data mining techniques like Finding dependencies, Cluster analysis and Classification analysis, it is the user who has to select the appropriate parameters for performing data analysis. As given in [LJLL, 01], an intelligent data mining system was developed in the Hong Kong Polytechnic University for a dispersed manufactured network. It works like other data mining systems for performing data analysis. It also doesn't have an integrated and automated approach with it.

Also most of the data mining systems are application dependent, no data mining system has the facility of selecting the parameters automatically from the given database for the given problem domain. By taken these points in view the new system has been proposed.

### **3. NEED OF THE PROPOSED SYSTEM**

The concept of domain knowledge is very important in data mining. In order to get the correct knowledge from the data mining system, the user must define the objective and specify the algorithms and its parameters exactly. If the data mining system produces large number of meaningful information by using a specialized data mining algorithm, it will take more time for the end-users to choose the appropriate knowledge. In some cases, even choosing the correct data mining algorithm involves more time and the domain knowledge. This proposed system deals about using intelligent agent based approach in solving data mining problems. If the user is using the system frequently the previous results also will be prompted when he enters in the system next time. Since the new system selects the algorithm and the parameter automatically, the user can perform different types of data analysis without wasting much time. Also the new system uses Intelligent agents, selecting the appropriate parameter for the given problem domain will be done by the system itself.

### **4. USING INTELLIGENT AGENTS**

Agents, special types of software applications, have become a very popular paradigm in computing in recent years. Some of the reasons for this popularity are their flexibility, modularity and general applicability to a wide range of problems. With the explosive growth of information source available on the Internet, and on the business, government, and scientific databases, it has become increasingly necessary for the users to utilize automated and intelligent tools to extract knowledge from them [Ays, 99]. Intelligent agents are an emerging technology that is making computer systems easier to use by allowing people to delegate work back to the computer. They help do things like find and filter information, customize views of information, and automate work. An Intelligent agent is software that assists people and acts on their behalf. Intelligent agents work by allowing people to delegate work that they could have done, to the agent software. [Gil, 97]

Using Intelligent Agents in data mining is quite interesting and also it will automate the process. Also automating the entire process may not be viable solution [AST, 00]. Because of the incorporation of multiple technologies to discover the knowledge, the new users will be able to interact with the system more efficiently. In this research work the following points are incorporated. Certain parameters like, user name, the database

name, field name, type of algorithm used are also stored in data base with the knowledge component. If the user is using the system many times for data analysis, the intelligent agent will choose the resultant database for giving the knowledge back to the users without mining one more time. Also, the featured system will prompt the results given to the user during the previous time. Whenever an update occurs, the intelligent agent program will caution the user and if the need arises, it will analyze only the updated part.

## 5. ALGORITHMS USED

The following recent data mining algorithms were used in this research work to design and implement the automated data mining using intelligent agents in object oriented databases

- a) Association Algorithms (APriori Algorithm, FP-Tree Growth Algorithm, Border Algorithm)
- b) Clustering Algorithms (BIRCH Algorithm, CURE Algorithm, CACTUS Algorithm)
- c) Classification Algorithms (K-Means / K-Medoid Algorithm, C4.5 Algorithm, CHAID Algorithm)

The new system uses any one of the above techniques to perform data analysis. If the user is using an algorithm initially, he/she may go for choosing another algorithm with different parameters also in the next time to get better / accurate results. As the preliminary step in the research, only few algorithms are considered for the development and implementation of the system. Many algorithms will be added in the future.

## 6. DATABASE USED

For implementing the above system a object data warehouse has been taken which has 1,50,000 records. The structure of the database is as follows

*“Age, Work-class, education, education-number, marital-status, occupation, relationship, race, sex, hours-per-week, native country, Income”*

The above object database is used to input to any one of the algorithms specified in the section vi. If the need arises the same database can be used with the different algorithm and parameters for performing sophisticated data analysis. From the above defined database, the associations between age and workclass, age and education, marital status and occupation, marital status and hours per week and many more can be analysed by using the

different techniques available. The other data mining techniques can also be used to perform the same with different parameters.

Object that share a common set of properties can be grouped into an object class. Each object is an instance of its class. Object classes can be organized into class / subclass hierarchies so that each class represents properties that are common to objects in that class. Object oriented databases are becoming increasingly popular in industry and applications. [JM, 01]. Because of this features the new system uses the object oriented concept.

## 7. FRAMEWORK OF THE NEW SYSTEM

The framework of the newly developed system is represented diagrammatically as follows. The automated data mining system gets the input from the given database (after the cleaning). Apart from getting the data alone, it gets the other information like, user information, and metadata. If the user is repeating the task, the new system will get the input from the storage also. Intelligent agent is used in the system to select the data mining technique and to get back the results from the storage. If the need arises, visualization tools may also be applied with the resultant database for pattern analysis.

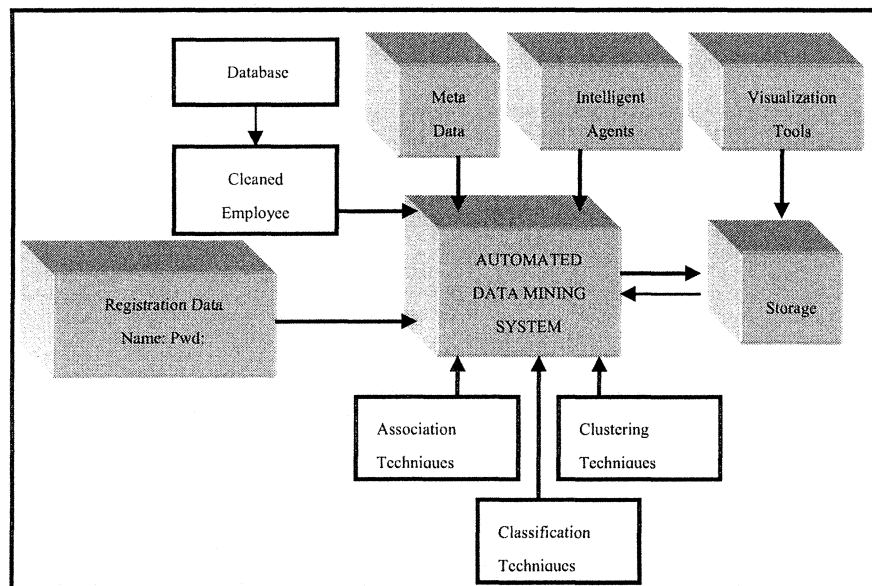


Figure 1: Framework of the newly developed automated data mining system

## 8. PERFORMANCE EVALUATION

The user needs only minimum time to perform the analysis in the newly developed system because of the using automated approach and intelligent agents. This in turn improves the overall efficiency of the system. Also the results produced by the new system will be better than the results produced by the existing system. The Prototype development of the proposed system is on the way using Visual Basic. The database is available in Oracle.

## 9. CONCLUSIONS

Data Mining is not the answer to all problems and sometimes it has been over emphasized [Thu, 00]. It is expensive to carry out the entire process and therefore has to be thought out clearly. Since the proposed system uses multiple technologies such as learning agents, data mining to discover the knowledge, the results produced by this system shall be more relevant and interesting than the results produced by any data mining system. Also the new users / end users will get better results quickly while operating this data mining system because of the easiness and user friendliness. Also, this research project is in the finishing level, the results and the future directions will be published shortly.

## REFERENCES

1. [AST, 00] Alex Berson, Stephen Smith, Kurt Thearling, "*Building Data Mining Applications for CRM*", Tata McGrawhill Publishers, 2000.
2. [Ays, 99] Ayse Yasemin Seydim, "*Intelligent Agents: A Data Mining Perspective*", Dept of Computer Science and Engineering, Southern Methodist University, Dallas, TX 75275, May 1999.
3. [Thu, 00] Bhavani Thuraisingam, "*Data Mining: Technologies, Techniques, Tools, and Trends*", CRC Press, 2000. Pages: 1, 4, 6, 93.
4. [Gil, 97] Don Gilbert, "*Intelligent Agents: The Right Information at the Right Time*" IBM Corporation, Research Triangle Park, NC USA, May, 1997
5. [LJLL, 01] H.C.W. Lau, Bing Jiang, W.B. Lee and K.H. Lau, "*Development of an intelligent data mining system for a dispersed manufacturing network*", Journal of Expert Systems, September 2001, Vol 18, No 4.
6. [JM, 01] Jaiwei Han & Miceline Kamber, "*Data Mining: Concepts and techniques*", Morgan Kaufmann Publishers, 2001, Pages 5-7.
7. [Puj, 01] Arun K Pujari (2001), "*Data Mining Techniques*", University Press, First Edition, Pages 79-173.