CHAPTER 6
CONCLUSION AND FUTURE SCOPE

6.1 Summary of Research Contributions

Due to the exponential growth of available information data accessing and mining systems are increasingly becoming critical to organizations, which wish to exploit operational and other available data to improve the quality of decision making. There is an urgent need for a new generation of automated tools to assist humans in extracting useful information (knowledge) from the rapidly growing volumes of digital data. A special, flexible software program, software agents used for automating the discovery of this information are required. The main objective of the thesis is to investigate and propose an approach for an automated data mining in object oriented data with the help of software agents. The goal is to empower non-expert users to achieve reasonable results with minimum effort. The automated data mining system must provide an interface to the non-experts and also should hide the mining concepts away from the users thus helping to bridge the conceptual gap usually associated with data mining. The following are the summary of the research contributions of the developed automated system.

i) The challenges of automated data mining system for object oriented databases are, first to find the attributes of relation assigned to relations. Design and development of an automated data mining system for object oriented databases is a challenging and relevant
problem. With the help of the user interface agent used in this research work the object oriented data is analyzed.

ii) The user interface agent used in this research work is very intuitive and provides assistance. The user interface agent observes and monitors the actions taken by the user in the interface and cooperates with the user in accomplishing some task in the application. The user interface agent analyzes the object oriented data set and it classifies them based on the following classification. It identifies the relationships between the objects, classes, and attributes. It identifies the importance of each of the attributes. The user interface agent also identifies the type of the attributes. The user interface agent is responsible for receiving user specifications and delivering back results. After the analyzes are carried out by the user interface agent based on the user’s high level goal or objectives or user’s hint the action takes place. The user interface agent identifies the attribute for the action.

iii) Vertical partitioning, partitions a class such that all attributes and methods of the class are closely related to each other. The analyzed and classified information is communicated to the partitioning agent. The Partitioning agent is an agent which helps in vertical partitioning. With the help of the partitioning processor the partitioning agent, partitions the object oriented data. The partitioning agent will decide
the number of partitions based on the relationship between the attributes, classes, objects, and methods. Once the object oriented data is partitioned, the ranking agent is called by the partitioning agent.

iv) The Ranking agent helps in ranking the attributes. Ranking is performed based on the query weight and scoring function. After the ranking of attributes the ranking agents communicates with the Data Mining agent. The highest ranked attributes are taken for clustering.

v) The data mining agent contains specific clustering algorithms implemented in house, that may have been installed within the environment. The Data mining agent decides which clustering algorithms are appropriate for the ranked data and it selects the appropriate clustering algorithm. Thus, the data mining agent is responsible for performing the actual data mining activity and generating the results. All the agents are able autonomously to work with less human intervention.

3.2 Major Strengths of the Approach

Considerable progress is being made in the areas of Automated data mining system for object oriented databases. Data Mining methods are highly technical and requires in-depth domain knowledge. However, by integrating data mining system with software agents empowers non-expert users to achieve reasonable results with minimum effort or with less human intervention.
Software agents used in this research work hides the complexity of the system thus helping the naïve users to perform complex data mining activities.

The major strengths of the approach are:

i) This research work deals with the design and development of an automated data mining system for object oriented data using software agents.

ii) Agents "learn" and evolve over time thereby improving the efficiency of the system.

iii) It has an open architecture and is scaleable; new agents can be introduced into the data mining system easily as long as they have a well defined interface.

iv) A new data mining system is designed and developed in this research work, to empower non-expert users to achieve reasonable results with minimum effort or with less human intervention.

v) Software agents hide the mining concepts away from the users thus helping to bridge the conceptual gap usually associated with data mining.

6.3 Future Scope

Data mining as a discipline has matured considerably, and there exists a multitude of scalable algorithms that transform oceans of bits in very large databases into interpretable patterns and predictive models. Automated data mining and modeling software gives a tool to perform analyses that otherwise would need to be handled by a highly trained researcher. Advances in technology
have expanded the areas of analysis that can be automated. Automated Data mining should incorporate those changes for further research. Some of the future research directions can be carried out under the following topics:

i) **Vertical partitioning of methods and inheritance**

This research work considers only simple classes with method and attributes. Complex objects with methods and inheritance have not dealt with. The current research work is attempting vertical partitioning of simple attributes and methods. A better partitioning solution can be obtained when physical characteristics of disk and access characteristics of transactions are incorporated in the cost function.

ii) **Ranking parameters or factors**

The ranking factors used in this research work are *query weight* and *scoring function*. There are various other ranking parameters to be considered for future work like *memory space, size of the object, etc.*. Ranking can also be carried out based on the memory space occupied by each attribute.

iii) **Implementing and automating all other data mining techniques**

At present this research work uses only clustering techniques for an object oriented data. The remaining data mining techniques such as association, prediction, classification, outliers etc can also be considered to develop a full fledged automated data mining system.
iv) Incorporation of all the Data Mining algorithms and automation

Although the data mining field is quite young, a number of algorithms had been proposed to discover various kinds of knowledge. In this research work only few commonly used clustering algorithms for object oriented data are implemented. All the other data mining algorithms can be implemented and automated for further research work which helps the user to mine the knowledge with less human intervention.

v) Multiple attributes

At present this research work deals with numeric, categorical and object oriented attributes. However knowledge can be mined from different attributes like time, spatial data etc. Exploiting different types of attributes in automated data mining system is an area with an enormous potential.

vi) Different kinds of data set

The present research work automated data mining system uses Campus Management System, Student data set of Karunya University. Partitioning, Ranking and clustering are carried out using the student data set. In future the developed automated data mining system can be made available in such a way to suit any kind of object oriented database
domains such as Telecommunication data, Health care data, Genomic data etc.

vii) **Different Partitioning Schemes for Object Oriented Data**

In this research work, emphasis has been given only to vertical partitioning schemes for object oriented databases. However, there are other partitioning schemes for object oriented databases. They are horizontal partitioning and path partitioning. These partitioning schemes can be incorporated in future into the developed automated data mining system. Based on the data set it should use any of the three partitioning schemes vertical or horizontal or path partitioning automatically by the use of software agents.

viii) **Visualization and data mining**

Visualization of database contents and data mining results may help users comprehend or appreciate mining results and redirect miners to search for promising patterns. Easy-to-use and easy-to-see visualization tools will be asset for automated data mining.

6.4 **Limitations**

- It has limited user friendliness, meaning users are expected to correctly set the goals or query when interacting with data mining system. Although, currently, the automated data mining system function can explore data only on clustering algorithms.
The data mining system has limited explanation and visualization facilities. It needs to better explain to its users the rationale behind its discovery of new knowledge, patterns, or trends.

It uses only clustering data mining techniques and clustering algorithms for different attributes.

6.5 Conclusions

Due to the exponential growth of available information data accessing, mining systems are increasingly becoming critical to organizations, which wish to exploit operational and other available data to improve the quality of decision making. Most current systems rely heavily on user’s interaction for inputs. The user must specify which data mining algorithm to use and all the parameters associated with the algorithm. The setting of these parameters requires a great level of expertise from the user. The user must do more work to decide which results are useful or interesting. Even if the required human resource is available, that person might still not be efficiently able to tackle the task on hand, due to the number of unknowns in the data to be mined.

A special, flexible software program, software agents are used for automating the discovery of this information. This research work deals with the design and development of an automated data mining system for object oriented data using software agents. A new data mining system is designed and developed in this research work, to empower non-expert users to achieve reasonable results with minimum effort or with less human intervention. Thus
the software agents hide the mining concepts away from the users thereby helping to bridge the conceptual gap usually associated with data mining.