© 2015 IJSRSET | Volume 1 | Issue 5 | Print ISSN : 2395-1990 | Online ISSN : 2394-4099
Themed Section: Engineering and Technology

A Framework for an Agent Based Computing using Data Mining Technique for Priceless Laptop Scheme of Tamilnadu Government

Dr. K. Shyamala*, S. Thilaka

PG & Research Department of Computer Science, Dr. Ambedkar Government Arts College, Chennai, Tamilnadu, India

ABSTRACT

Predicting the performance (usefulness) of any welfare scheme will be more useful for the government to make decision whether to continue or to drop out the schemes in the future. To attain this objective, the concept of agents and data mining can be used. Data mining techniques are used to discover models or patterns of data and are much helpful in the process of decision making. Intelligent agents represent an important opportunity to optimize knowledge management. The data used for this study is collected, evaluated and analysed by means of agent paradigm along with the clustering techniques of data mining technology.

Keywords: Agent Based Computing, Data Mining, Software Agents, Agent Mining, Clustering.

I. INTRODUCTION

Education in every sense is one of the fundamental factors of development. Education enriches the people's understanding of themselves and of the world. It improves the quality of their lives and leads to broad social benefits to individuals and society. Recognizing the significance of education in the development process and the economic imperative of "Quality of Education for all", the state Government has been encouraging students through various welfare schemes. One of the best schemes of state Government, which improves the technical skills of students', is "Providing free laptops to school and college students". This paper is designed to analyse the usefulness, activities, feel and the suggestions of students regarding the free laptops, based on their data [1]. The clustering technique of data mining technology is used for this study along with the agent paradigm. Detecting the interesting patterns and finding out the appropriate knowledge from the output is also an interesting issue in data mining. This research work addresses this issue with the help of software agents. The software (intelligent) agents help in detecting the clusters automatically. It also assists in viewing the cluster results graphically. Multidimensional visualization is used to view the results in a more meaningful way [2].

1.1 Intelligent Agents and Data Mining

Agents, i.e. special types of software applications, have become increasingly popular in computing world in recent years. Some of the reasons for this popularity are their flexibility, modularity and general applicability to a wide range of problems (data filtering and analysis, information brokering, condition monitoring and alarm generation, workflow management, personal assistance, simulation and gaming) [3].

Because of the explosive development of information source available on the Internet and on the business, government, and scientific databases, it has become quite necessary for the users to utilize automated and intelligent tools to extract knowledge from them [4]. Intelligent agents can help making computer systems easier to use, enable finding and filtering information, customizing views of information and automating 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 [2].

On the other hand, data mining is the process of posing queries and extracting useful information, patterns and trends previously unknown from large quantities of data

[5]. Data mining is also a multidisciplinary field, working in areas that include artificial intelligence, machine learning, neural networks, pattern recognition, knowledge-based systems, information retrieval, high performance computing, and data visualization [6].

The concept of 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 with minimum effort. If the data mining system produces large number of meaningful information by using a specialized data mining algorithm (association, clustering, decision trees etc.), it will take more time for the end- users to choose the appropriate knowledge for the problem discussed. Even choosing the correct data mining algorithm involves more time for the system. A solution for this problem could be an intelligent system based on agents. Data mining and intelligent agents can make a common front to help people in the decision making process, to elaborate decisional models and take good decision in real time" [3].

1.2 Cluster analysis

Clustering is the process of making a group of abstract objects into classes of similar objects. Cluster is a group of objects that belongs to the same class. In other words, similar objects are grouped in one cluster and dissimilar objects are grouped in another cluster.

- A cluster of data objects can be treated as one group.
- While doing cluster analysis, we first partition the set of data into groups based on data similarity and then assign the labels to the groups.
- The main advantage of clustering over classification is that, it is adaptable to changes and helps single out useful features that distinguish different groups.
- Clustering analysis is broadly used in many applications such as market research, pattern recognition, **data analysis**, and image processing.

1.2.1 Clustering Methods

Clustering methods can be classified into the following categories

- Partitioning Method
- Hierarchical Method
- Density-based Method
- Grid-Based Method
- Model-Based Method
- Constraint-based Method

Of all the above methods, the method used in this research work is density based method.

Density-based Method

This method is based on the notion of density. The basic idea is to continue growing the given cluster as long as the density in the neighbourhood exceeds some threshold, i.e., for each data point within a given cluster, the radius of a given cluster has to contain at least a minimum number of points.

II. LITERATURE REVIEW

Over the past decade, rapid technological advances have sparked interest in utilizing laptops as an instruction tool to improve student learning. As per the article of Deccan chronicle, the survey done by the private company has revealed that eight out of 10 students seemed to be happy with the quality of the free laptop provided by the Tamil Nadu Government. The survey also state that, "free laptops improves technical skills of students". The Indian higher education system is presently facing several challenges. Various welfare schemes are to be implemented by the state Government and it has to be continued till the people become self-sufficient. The clearest areas in which the laptop program has had a positive impact are in improving computer skills, increasing access to educational resources, boosting student motivation and interest in school/college and enhancing interaction among students and teachers. The quality of the schemes and whether the current products meet the user needs should also determine by the study [1].

The developed automated system gets the input values form the user and chooses the appropriate data mining techniques with required parameters by using intelligent agents [2]. Ayse Yasemin SEYDIM's [7] explained more on agents, the special types of software applications, has become a very popular paradigm in computing in recent years. The author states that, the agent based studies can be implemented for clustering, classification, and summarization. Some of the reasons where agents are more flexibility, modularity and

general applicability to a wide range of problems. Recent increase in agent-based applications is also because of the technological developments in distributed computing, robotics and the emergence of object-oriented programming paradigms. Advances in distributed computing technologies have given rise to use of agents that can model distributed problem solving.

Vuda Sreenivasa Rao [8] explained communications among the agents with in multi-agent system. According to the author, multi-agent system often deals with complex applications that required to solve the existing problem during data mining process in distributed system with individual and collective behaviors of the agents depends on the observed data from distributed system. Based on this concept, an integration of multi-agent system with data mining is incorporated and it also defines how multiple agents are communicated with respect to specific applications. Declaration of different agents with respect to specific task and communication behavior among agents is considered in this research work to meet the user requirements.

David Meunier and Helene Paugam-Moisy [DH, 9] focused on Girvan & Newman (GN) method which was created by Girvan & Newman to determine the clusters in a given undirected graph, without any restriction on the data size and number of clusters. GN method is based twopass algorithm where the first pass removes the edges with highest edge – betweenness centrality and cluster building in the second pass. Author proposed an another method based on GN method instead of undirected graph to directed graph in the name of arc-linked cluster detection method with neural networks concepts to achieves a narrower and higher modularity peak than GN method. The resultant value is more pertinent optimal set of clusters. The time complexity of Arclinked based is higher than GN method. The concept discussed by these authors is considered in this research work during cluster detection phases after cluster formation to detect the quality of cluster by means of attribute selection from the given database.

Wout Dullaert, Tijs Neutens, Greet Vanden Berghe [WTG, 10] implements an intelligent agent based communication for particular platform. Intelligent agents are used in the form of high potential output such as increase cost efficiency, better service and safety communication among the agents. They are also autonomous, communicative and intelligent. Author also proposed real-time decision is also possible with the presence of intelligent agent. Agents are used to overcome the quality, reliable service, trust concerns and confidentiality during the exchange process. Agent technology is used for automated transport process. In this research work, intelligent agent-based concept is considered for cluster formation and cluster visualizations.

Nigel Robinson, Mary Shapcott [11] proposed a visualization aids (beyond charts and graphs) by consider virtual data mining environment and data set as liquid data. The author focused was based on limitations of the existing visualization methods and problems faced by the user while visualization. As innovation, the modified visualization result is in the form of 3D game representation so that all users can easy to understand without having domain knowledge. The minimization of user difficulties is taken into account in our research process during visualization. In this paper, automation of the process is not yet considered for result oriented. Because of this, it is difficult to find the results by rare user. From the above procedure, concept of visualization is taken into account as a part of this research work in which an automated data mining system is implemented to sense the user behavior for visualization so as to minimize the difficulties faced by the user during graphical representation of the results.

Dr. Ping Chen, Dr. Chenyi Hu, Dr. Heloise Lynn, and Yves Simon [12] discuss about visualization of high dimensional data which is very important in data analysts due to its visual nature. They also proposed a method to visualize large amount of high dimensional data in a 3-D space by dividing the high dimension data into several groups of lower dimensional data first. Then, different icons are used to represent different groups. A glyph-based technique is used to represent different set of data in the form of various color icons like line, point, polygon, etc. The visualization of high dimensional data using Glyphs – based techniques is considered in this research work for visualization of multi-dimensional data based on user expectation.

III. METHODS AND MATERIAL

As mentioned above, Data mining techniques are used to discover models or patterns of data and are much helpful

in the process of decision making. In the proposed system, the density based clustering technique of data mining model has been used to study the factors related to the usefulness of Priceless Government Laptops scheme. To expertise this process, the role of agents comes into picture. The proposed framework [13] has the following agents…

A. User Interface Agent

The user interface agent interacts with the user in assisting him / her to perform data analysis and data mining activities. The user can provide a general description of the problem at hand in terms of high level goals and objectives, or provide specific details about the data analysis or mining tasks to be performed. [14] Given this, three interesting questions arise.

- What aspects of the user's behaviour are useful to capture?
- What information such behaviour gives us on the actual intentions and preferences of the user?
- Finally, what use can be made of any such information captured?

The user interface agent provides the solution to the above questions in the following way: Based on the user behaviour the productive attributes are mined from user agent by data mining agent where the actual mining process takes place. User interface agent analyses all the information with respect to the user behaviors and transfer the control to data mining agent for further process.

B. Data mining Agent

A data mining agent is a pseudo-intelligent computer program designed to ferret out specific types of data, along with identifying patterns among those data types. These agents are typically used to detect trends in data, alerting organizations to paradigm shifts so effective strategies can be implemented to either take advantage of or minimize the damage from alterations in trends. In addition to reading patterns, data mining agents can also "pull" or "retrieve" relevant data from databases, alerting end-users to the presence of selected information [15].

The data mining agent is responsible for performing the actual data mining activity and generating the results. The specific data mining methods and algorithms are implemented by the data mining agent. It has two major roles: cluster formation and cluster detection based on the data.

- **Cluster formation**

Cluster formation is defined as grouping of objects that are similar to one another within the same cluster and are dissimilar to the objects with other clusters. Partitioning approach is considered to cluster different types of attributes, numeric and categorical data. Clustering are also done depends upon data set, data size, and data types. Data mining agent will choose the appropriate clustering algorithm for better cluster formations.

- **Cluster detection**

Data mining agent is used here to analyze the formulated clusters quality based on quality parameters. Generally, clustering algorithms will produce clusters, based upon input data. But, all clusters are not good cluster. Intelligent agent in automated system is used to find the good clusters among various clusters.

C. Visualization Agent

To evaluate the effectiveness of visualization techniques, the way in which they assist and complement the data mining process need to be understood. The relationship between data mining and visualization process can be explained with few conditions visualization of data set can be defined as combination of various methods or user priority to an approach of selecting and indicating what patterns should be displayed. It is one of the interesting techniques to establish which patters are better than the one to enhance visualization techniques. Visualization will provides more added advantage to the user for easy understanding and also increases the power of understanding the end results.

The visualization agent is uses to generate various reports based on the cluster results. The visualization might be 1D, 2D or 3D based on the type of cluster nature (numerical or categorical). Generally, major difficult task in data mining is report generation, that to it should be understand able by all the users. Manually it

is possible to create reports but it is hard to justify the report at all times. In this work, visualization agent will coordinate with data mining agent, based on that it will identify suitable visual method for each specific cluster [13].

From these aspects, a framework is developed for data analysis, cluster detection and visualization for better decision making by means of intelligent agent technology. The methodology adopted in this research work is described as follows…

Proposed methodology

Step 1: Start
Step 2: Pre-process the data
Step 3: Design a suitable procedure for intelligent agents to analyze the data
Step 4: Selection of appropriate clustering technique
Step 5: Selection of appropriate visualization tool
Step 6: Multi-dimensional visualization of results
Step 7: Stop

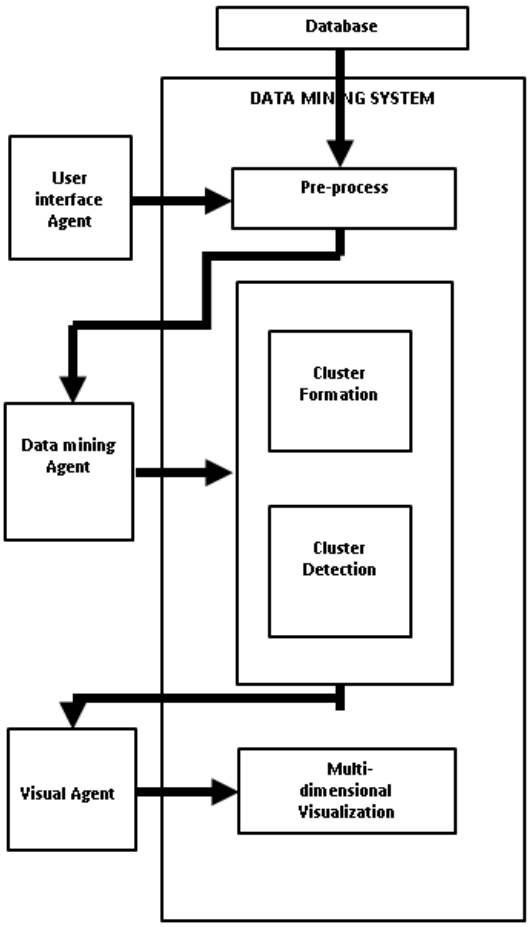

Figure 1: Proposed framework for agent based computing using data mining technique

The newly developed framework is represented diagrammatically is shown in Figure 4.1. From this framework, it is clear that the data mining system gets the data from the database for its processing.

Apart from getting data from the given database, it also gets the user specifications through the user interface agent. The data mining agent chooses the appropriate algorithm for the pre-processed data received from the user interface agent. The algorithm used here is the density based clustering algorithm. The data mining agent is responsible for the cluster formation and detection for the datasets available. Once the selection of appropriate clustering algorithm is over, the next step is to select the suitable visualization tool. The visualization agent is responsible for displaying the results in an easy understandable format, which can be used to portray the exact results of the survey. The intelligent agents are used in automated process of attribute selection, ranking process, cluster formation, cluster detection and visualization.

The user interface agent is responsible for receiving user specifications and delivering results to the data mining agent. The data mining agent mine the result based on the parameters given by user interface agent to perform cluster formation and cluster detection. Clustering algorithms are used to formulate a new cluster, based on the user interface agent with respect to specific user profile. Then cluster detection techniques are used to detect the cluster quality for further process. The quality of cluster is identified by various parameters with help of intelligent agent. As result, the best cluster will be discovered from the known knowledge. The visual agent is used to visualize the identified cluster depends on the nature of data within the detected cluster. After completion of cluster formation and cluster detection, it transfers the result to decision system where the visual agent find suitable representation tools based on the cluster nature. Finally, the end result is visualized in terms of 1D, 2D or 3D by visual agent. Thus, the entire process is monitored as well as executed by automated intelligent system based on the user profile. This makes the less domain knowledge users more convenient and understands ability.

In this research work different types of agents are used to perform the operations on behalf of user so that, the data mining result will be productive and knowledgeable

for less domain knowledge user. Agents used in this framework are for reliable communication, cooperation among the agents, and finally coordination among the other agents within the system to perform some specific tasks [13].

4.1 Database considered

Education acts an integrative force in society, imparting values that posters social position and national identity. Recognizing the importance of education the state Government has placed an unprecedented focus in expansion of education, significantly improving the quality of education imparted and ensuring that educational opportunities are available to all segments of the society [1]. To enhance the knowledge of students, the scheme of free laptops is introduced. So far, various surveys are conducted in various departments such as commerce and economics to analyze the quality of this scheme. Implementing this in the field of computer science is the first of its kind. The concept of data mining is used to perform this analysis. WEKA 3.7, one of the prominent tools of data mining is used for this study. It provides a collection of machine learning algorithms for data mining tasks. WEKA contains tools for data pre-processing, classification, association and clustering rules and visualization. The DBSCAN algorithm of clustering technique is used to carry out the study.

To perform the analysis the test data was collected from Dr. Ambedkar Government Arts College, Chennai from various courses. All students of all the departments are encouraged to participate in the survey. At least two percent of all the classes are covered. Nearly 500 students are participated in the survey. Out of these participants 53% are male students and 47% are female students. It contains nearly 27 attributes of various data types. Some of the attributes are as follows: NAME, GENDER, CLASS, AVAIL_BF_LAP, POSSESSION, USG_PURPOSE, SUGGESTION, DURATION, STUDENTS_FEEL, ADD_FEATURES etc.

IV. RESULTS AND DISCUSSION

There were nearly 28 attributes were analyzed. Some of the important findings of this study are presented below…

A. Specific purpose of usage

The specific purpose of using laptops by the students are classified as follows: 19% of the students use their laptop as study material, 16% use them for playing games, a single percent use for hearing songs and 10% uses them to watch movies and the major category, 54% of them use them for all the activities they are like to do such as playing games, watching movies, hearing songs, taking study materials etc.

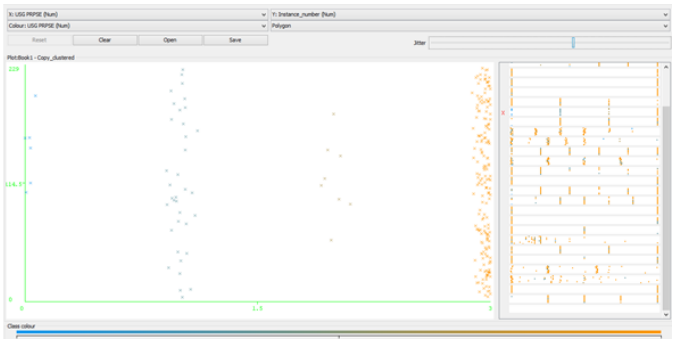

Figure 2: Specific purpose of usage

B. Students' Feel

The survey included an open ended question asking students about their feel when they carry laptop with them. Few more questions are also raised to understand the pulse of the students. First their opinion about giving laptop to students other than computer science students is questioned. 86% of them feel it is useful, 13% reported that they have no idea and the balance 1% feels it is not needed.

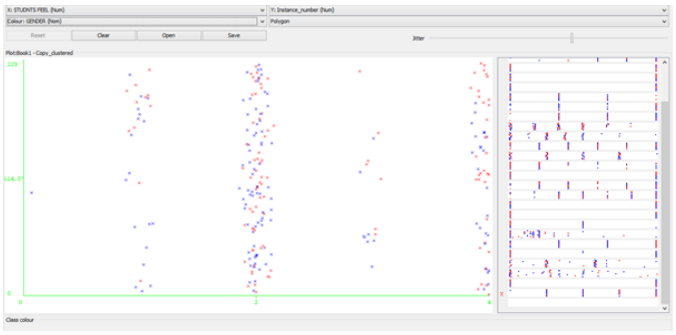

Figure 3: Students' Feel

Next the likeness about selling their laptop is analyzed. 60% of the survey respondents reported they don't like to sell their laptop. Third the students are asked whether the laptop has become an essential part of their life. Hopefully, 71% of them reported as it is essential. Now considering the feel 69% of the students feel proud and

happy with the Priceless Government laptop scheme provided by the Government of Tamil Nadu.

C. Suggestions

Students' stated many suggestions regarding the model of the laptop. The major category 77% of the students reported that they don't have any suggestions in improving the model. The next set of students suggested to provide the laptop with CD- drive and the touch model (each 5%). The rest of the students made various other suggestions. The suggestions made by them are visualized in the form of clusters as follows…

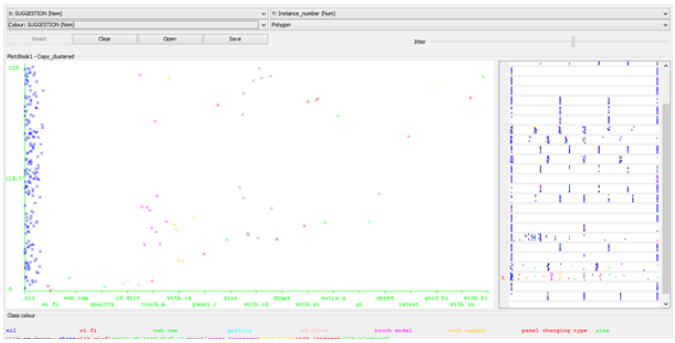

Figure 4: Suggestions provided by the students

V. CONCLUSION

Data mining is the procedure of mining knowledge from data, which can be widely used in the process of decision making. An agent is a computer system located in some environment, and is capable of autonomous action in this environment to meet its design objectives. Multi agent systems are systems made up of multiple interacting agents. In this research work, the mining process is carried out by different intelligent agents with the help of the data collected. The results are analyzed, evaluated and visualized by forming and detecting the clusters using user interface, data mining and visualization agents.

The major aim of this research work is to predict the usefulness of Priceless Government laptop scheme of Tamil Nadu. This study finds that there is convincing evidence that the laptop program has contributed much to improvement of students' performance. The laptops have given to Government / Government aided school and college students, who are basically from a poor family. So, it is a gift for those who cannot afford the price of owing laptops. The laptop provided by the Government is not only useful for that one particular person but also to the whole family, either the sister or brother of him/her can make use of that laptop for their school projects and also for their further studies. It improves the technical skills of both school and college students. Though it has certain disadvantages (distraction of students in the classroom, using laptop for non-academic purposes more than academic activities like playing games and watching movies), the scheme's objective of enhancing students' knowledge is achieved. As per as our survey, we analyzed that the issue of Priceless Government laptop scheme is truly a boon to all students.

VI. FUTURE WORK

The future scope of this work is to carry out the analysis by using various other suitable methods and by using other data mining techniques. This model can be extended by implementing the proposed agent based computing framework using the various agent tools available. Its scope for future work can also be extended in the part of gathering data by increasing the number of instances from different colleges.

The data collected in this work is collected from the students' point of view and this can be extended by collecting data from teachers' and parents' point of view. This research work focussed only on the scheme implemented in Tamil Nadu. This can be extended by making comparative study between laptops issued by different state governments.

VII. REFERENCES

[1] S. Thilaka and K.Shyamala," An analytical study on the usefulness of priceless government laptops", International Journal OF Engineering Sciences & Management Research, ISSN 2349-6193, Impact Factor (PIF): 2.243

[2] Don Gilbert, "Intelligent Agents: The Right Information at the Right Time", IBM Corporation, Research Triangle Park, NC USA, May, 1997.

[3] Irina Tudor and liviu Ionita," Intelligent Agents as Data Mining Techniques Used in Academic Environment".

[4] Seydium, A.Y. (1999): "Intelligent Agents: A Data Mining Perspective", Dept. of Computer Science and

Engineering, Southern Methodist University, Dallas, TX 75275.

[5] Thuraisingam, B. (2000): "Data Mining: Technologies, Techniques, Tools, and Trends", CRC Press, 4- 6.

[6] Han, J., Kamber, M. (2001): "Data Mining: Concepts and techniques", Morgan Kaufmann Publishers, 5-7.

[7] Ayse Yasemin SEYDIM "Intelligent Agents: A Data Mining Perspective" Southern Methodist University, Dallas, 1999.

[8] Vuda Sreenivasa Rao," Multi Agent-Based Distributed Data Mining: An Over View", International Journal of Reviews in Computing, pp 83 – 92, ISSN: 2076-3328, E-ISSN: 2076-3336.

[9] David Meunier, Hélène Paugam-moisy," Cluster detection algorithm in neural networks",Proceedings of 14th European Symposium on Artificial Neural Networks (ESANN 2006), M.Verleysen (ed.), Pp 19-24, Bruges, Belgium.

[10] Wout Dullaert , Tijs Neutens, Greet Vanden Berghe , Tijs Vermeulen , Bert Vernimmen , Frank Witlox, MamMoeT: "An intelligent agent-based communication support platform for multimodalTransport", Expert Systems with Applications 36 (2009) Pp10280–10287.

[11] Nigel Robinson, Mary shapcott, "Data Mining Information Visualization Beyond Charts and Graphs", Proceedings of the sixth International Conference on information Visualization (IV ’02) 2002 IEEE.

[12] Dr. Ping chen, Dr. chenyi Hu, Dr. Heloise Lynn, and Yves Simon, "Visualizing Multi- Dimensional Data", www. Research.acxiom.com, 2002.

[13] R. Jayabrabu, Dr. V. Saravanan, Prof. K. Vivekanandan, "A framework: cluster detection and multidimensional visualization of automated data mining using intelligent agents", International Journal of Artificial Intelligence & Applications (IJAIA), Vol.3, No.1, January 2012.

[14] http://www.cs.upc.edu/~bejar/aia/aia-web/green97software.pdf

[15] http://www.wisegeek.com/what-is-a-data-mining-agent.htm