Volume 3, Issue 12, December 2013 ISSN: 2277 128X

International Journal of Advanced Research in Computer Science and Software Engineering

Research Paper

Available online at: www.ijarcsse.com

Comparative Examination of Data Mining Tools in Representative Based Systems

Vishnu Kumar, Sunil Sharma, Mithlesh Bekadia
(Research Scholar)
Jagannath University, Chaksu,
Jaipur, India

Prof. A.K. Tiwari
(Director)
Wilfred Institute of Information Technology
St. Wilfred's College, Jaipur, India

Abstract: The Worldwide technological advancement has brought in an extensive transform in adoption and consumption of open source tools. Since, a large amount of the organizations across the globe convention with a large amount of data to be updated online and dealings are made every second, managing, mining and giving out this dynamic data is very complex. The Successful achievement of the data mining technique requires a cautious assessment of the variety of tools and algorithms available to mining experts. This paper provides a relative study of open source data mining tools obtainable to the professionals. The Parameters influencing the preference of apt tools in addition to the real time challenges are discussed. However, it's glowing established that Representative aid in improving the presentation of data mining tools. This article provides the information on Representative-based framework for data preprocessing with execution particulars for the development of enhanced tool in the market. An incorporation of open source data mining tools with Representative simulation enable one to implement an effective data pre processing architecture thereby providing robust capabilities of the application which can be upgraded using a minimum of pre planning requirement from the application developer.

Key Words: Representative, Data mining, Representative replication, Representative based framework

I. Introduction to data Mining

Data mining, *the extraction of hidden predictive information from large databases*, is a powerful new technology with great potential to help companies focus on the most important information in their data warehouses. Data mining tools predict future trends and behaviors, allowing businesses to make proactive, knowledge-driven decisions. The automated, prospective analyses offered by data mining move beyond the analyses of past events provided by retrospective tools typical of decision support systems. Data mining tools can answer business questions that traditionally were too time consuming to resolve. They scour databases for hidden patterns, finding predictive information that experts may miss because it lies outside their expectations.

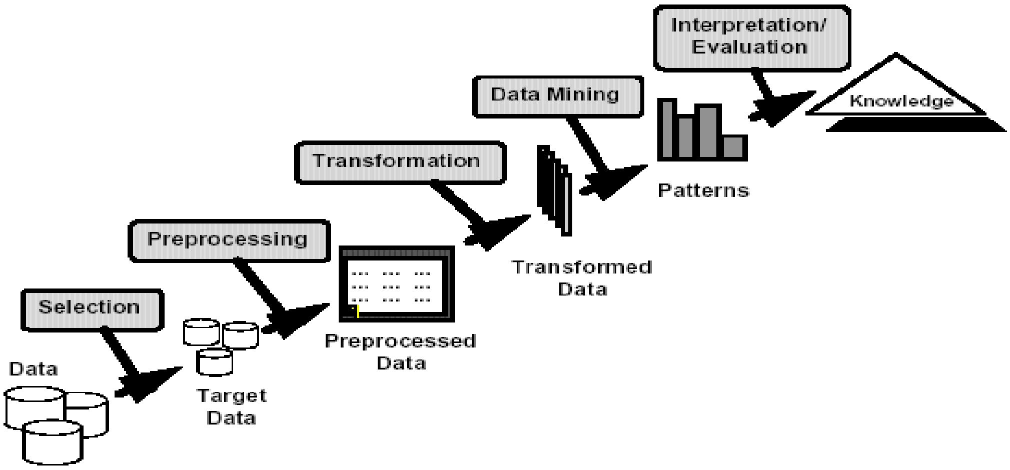

Figure 1 Showing Data Mining Process

Databases process comprises of a few steps leading from raw data collections to some form of new knowledge. The iterative process consists of the following steps are given below:

- **Data cleaning**: also known as data cleansing, it's a phase in which noise data and irrelevant data are removed from the collection.
- **Data integration**: the multiple data sources, often heterogeneous, may be combined in a common source.
- **Data selection**: the data relevant to the analysis is decided on and retrieved from the data collection.

- **Data transformation**: it's also known as data consolidation, it is a phase in which the selected data is transformed into forms appropriate for the mining procedure.
- **Data mining**: it's the crucial step in which clever techniques are applied to extract patterns potentially useful.
- **Pattern evaluation**: it's strictly interesting patterns representing knowledge are identified based on given measures.
- **Knowledge representation**: it's the final phase in which the discovered knowledge is visually represented to the user. This essential step uses visualization techniques to help users understand and interpret the data mining results.

It is common to combine some of these steps together. For instance, *data cleaning* and *data integration* can be performed together as a pre-processing phase to generate a data warehouse. *Data selection* and *data transformation* can also be combined where the consolidation of the data is the result of the selection, or, as for the case of data warehouses, the selection is done on transformed data. The recent advance in technology allows organizations to accumulate complex data in various locations and also allows data updating in every second. The main challenge of any organization is to effectively manage large updated data online. One of the influencing features that have constantly hindered the Data Stream Management System is its inability to simultaneously query both live and archival data

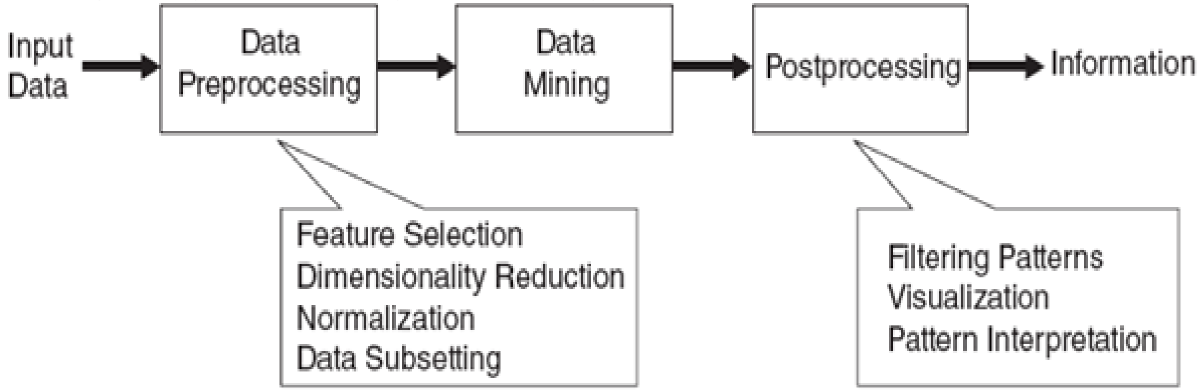

Figure 2 Showing Process of Knowledge Discovery in Database

1. Data Mining is the process that helps to make use of the data in various databases and find new patterns in it. Data mining model is successfully able to solve emergent problems such as the discovery of patterns and knowledge in uncertain, high-frequency, organizational, or behavioral data, including data generated and stored in various locations systems.
2. Applying data mining to acquire updated knowledge leads to complexity in data processing, managing and mining Commonly used open source data mining tools provide us with user friendly interface for data analysis and its main features are Handling Complicated Problem, Discovering Unknown Patterns, Skill Required in Working with the Tool, Scalability, Data mining tools should be capable of handling large amount of datasets and Cost.
3. According to Computer system environment for data mining tools require a client server environment.
4. The data mining tools has to be analyzed based on parameters like Product track record, Vendor Viability, Breadth of Data Mining algorithms, Compatibility with a specific computer environment, Ease of use and ability to handle large databases Representative based systems are the outstanding approach to overcome the drawbacks cited above. In recent years, Representative has become popular paradigm in computing because of its autonomous, flexible, adaptive and intelligent characteristics.
5. The intelligent Representative do the work with human intelligence but may not behave like human beings. The Representative which is involved in processing of mining performs productive task, retrieves useful knowledge with less noise and reduced processing time when compared to the normal mining tools
6. The Potential features of Representative based data mining tools include are following.
   - To Propose the processing techniques most suitable to the data
   - To Preprocess incoming new data according to user profile
   - To Share mining experience and Suggesting possible knowledge that can be extracted from the data with the help of the experience shared Suitable for novice user.

II. Obtainable Representative Replication Tools

The Java-based autonomous Representative developed by IBM, which make available the basic capabilities required for mobility and has a globally unique name. A travel schedule is used to specify the destinations to which the Representative must travel and what actions it must take at each location. In order for an aglet to run on a particular system, the target system must be running an aglet host application which provides a platform-neutral execution environment for the aglet. The aglet workbench includes a configurable Java security manager. The Aglets can communicate using a whiteboard that allows Representative to collaborate and share information asynchronously. Synchronous and asynchronous message passing is also supported for aglet communication. Aglets are streamed using standard Java serialization or externalization. A network Representative class loader is supplied which allows an aglet's byte code stream and state to travel across a network. The FTP Software Representative Technology is Java-based software designed to manage heterogeneous networks across the Internet using Representative technology. The Representative are autonomous and mobile, and can move to any system in the network which has Representative Responder installed. As the Representative moves from system to system, its tasks may change, depending on the environment of the system it is visiting.

The Representative can interrelate with other Representative or with the user, as desirable But FTP Representative do not require any user interaction based on push technology, they can move from system to system, respond to events, and perform tasks according to criteria predefined by the user. Representative Manager is responsible for launching the Representative. The explorer, from Object Space, Representative-enhanced Object Request Broker (ORB) coded in Java. An ORB provides the capability to create objects on a remote system and invoke methods on those objects. Voyager augments the traditional ORB with Representative capabilities. Voyager Representative has mobility and autonomy which Is provided in the base class, Representative. Representative can move itself from one location to another and can leave behind a forwarding address with a secretary so that future messages can be forwarded to its new location. Specialized Representative, called Messengers, are used to deliver messages. Messages can be synchronous, one-way similar to asynchronous, or future, which are asynchronous but return a placeholder that can be used to retrieve a return value at a later time. Representative's itinerary instructs the Representative as to what operations it needs to perform at each location. Voyager has a security manager which can be used to restrict the operations Representative can perform.

III. Architecture Development

The appearance of online transactions, World Wide Web, data streams. The large amount of data is created and is being stored in databases, which is one of the reasons for complexity of data management. apprentice users may find it difficult to determine the best technique for pre-processing their data to perform data mining. Hence, experts are required to identify the best-suited pre-processing technique. At present, neither data mining tools are efficient in handling dynamic and complex data nor do the users have sufficient knowledge in pre processing the data in terms of data mining domain. However, with the integration of Representative based data mining system, the Representative determines the technique and the parameters that provide the best model for good decision making. Since, the current mining tools are domain specific, this research focused us to propose a generic architecture that can pre-process data using Representative of any domain of application. In addition, even a novice user can use the proposed architecture

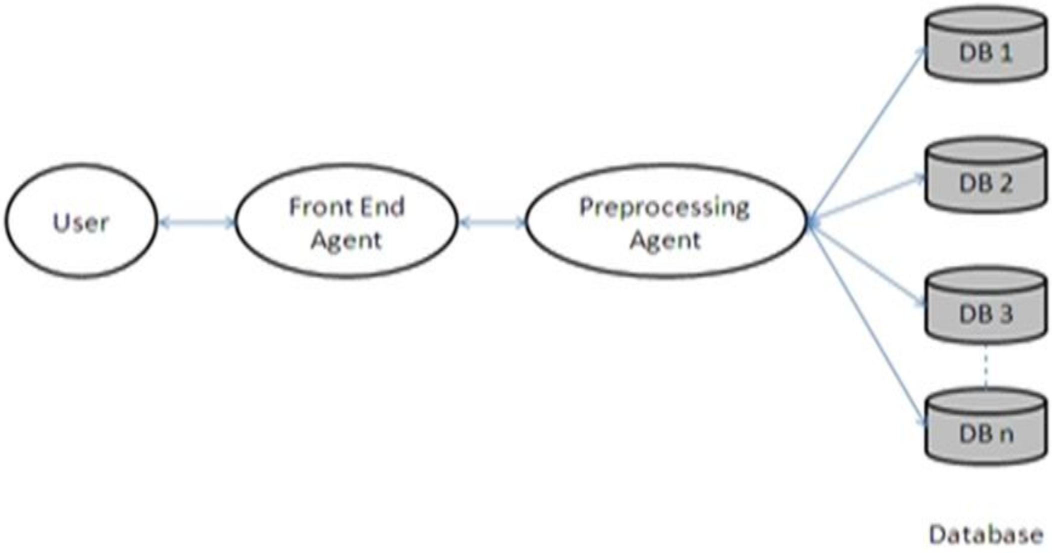

Figure 3: Showing Proposed Pre-processing Architecture.

The proposed architecture in figure 3 is the basic model of Pre-processing architecture. The framework is designed with five major Representatives which include User Interface Representative (UI Representative), Coordinator Representative, Clean Representative, Transformation Representative and Reduction Representative. Figure 2 depicts the aforementioned design of pre-processing architecture. The responsibilities and capabilities are specific of each Representative. The User Interface Representative provides the user interface with the system. It provides solution analysis autonomously and helps the user to give queries according to the requirement of user. Coordinator Representative is responsible for the coordination of all the tasks that is performed in the system. It determines the pre-processing method to be used based on the data mining task given by the user which is generated according to the Meta knowledge which the Representative maintains. It has access to the data repository that can be updated dynamically and provides data to the other Representative. Coordinator Representative also provides adaptive profiling user data and checks to identify the data types and attributes in the database. Its identifies the problems the data has and save the knowledge and corresponding preprocessing technique that is best for it. Clean Representative handles the missing and noisy data using the techniques in the data dictionary. Transformation Representative is used to transform the data into appropriate forms for mining. The role of reduction Representative is to reduce the size of the data by using preferred discretization techniques.

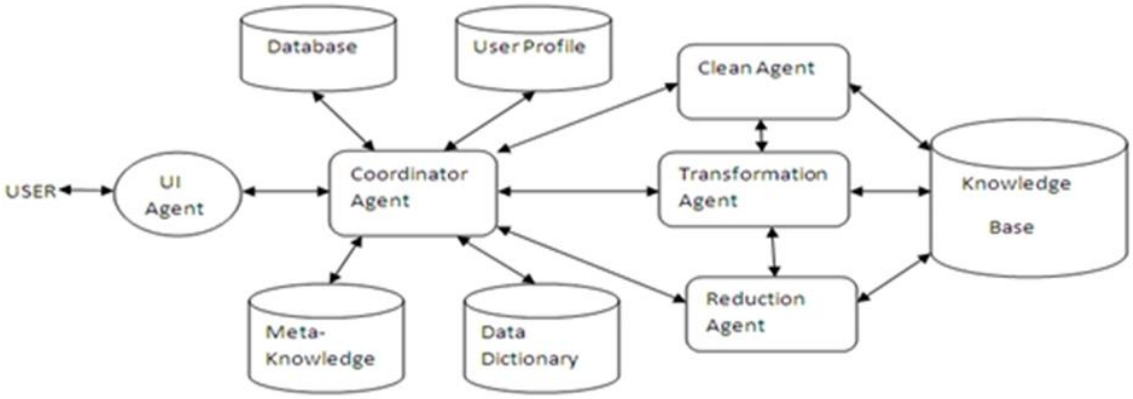

Figure 4: Showing Pre-processing Architecture

IV. Performance of the Representative Based Preprocessing Architecture

The proposed architecture model is a client server model with a basic Model View Controller Architecture. The exceeding proposed architecture can be implemented in similar ways like implementing both Representative and request or to modify the existing code of an application to enable the necessary communication. The preprocessing methods are all available in various open source data mining tools and open source Representative simulation tools.

The Representative framework which we proposed will have the following requirements

i. Capability to add intelligence to applications.
ii. Intelligent Representative framework is practical to solve real-world problems.
iii. Architecture must be flexible enough to support any applications.
iv. Intelligent Representative increases the functionality of the application and can communicate with each other and other applications.
v. The Representative can call the shots and monitor and drive the applications.

Useful requirement of Representative Framework

The functionality of Representative Framework for all time contain the following condition

- It must be easy to add an intelligent Representative to an existing application.
- A graphical construction tool must be available to compose Representative out of other Java components and other Representative.
- The Representative must support a relatively sophisticated event processing capability. Representative will need to handle events from the outside world, other Representative, and signal events to outside applications.
- Domain knowledge can be added to Representative using if-then rules, and support forward and the direction of the back rule-based processing with sensors and effectors.
- The Representative must be able to learn to do classification, clustering, and prediction using learning algorithms.
- Multi Representative Applications must be supported using a KQML-like message protocol.

The Representative should be determined. That is, once Representative is constructed, there must be a way to save it in a file and reload its state at a later time. Due to the availability of open source data mining tools and Representative simulation tools for the implementation of the architecture, it is now possible to append Representative to the existing application, thereby extending the basic capabilities of the application which requires a minimum of pre planning of the application developer.

V. Conclusion

State of the art advancement in technology enables organizations to store, process and update huge and complex data dynamically. The development and application of data mining techniques requires the use of right choice of software tools. Additional, recent data mining tools are expensive to get the updated knowledge model. This paper provides information on two aspects of data mining tools namely to elucidate the comparative analysis of various open source data mining tools and to put forth the challenges which exist in the existing data mining tools. However, Representative is known to aid in improving the presentation of data mining tools. This paper has therefore projected to integrate existing data-mining tool with Representative in order to execute an effective data preprocessing architecture. The useful specifications elucidate here enable the application developer to correctly analyze, assess and develop the data pre processing tool for improved data management.

References

1. Mohammed A Qadeer, Nadeem Akhtar, Faraz Khan ***"Comparison of Tools for Data Mining and Retrieval in High Volume Data Stream"***, 2013.
2. A. Mohtar. ***"MultiRepresentative Approach to Stock Price Prediction".*** University Kebangsaan Malaysia. 2013.
3. Longbing Cao, Vladimir Gorodetsky, Pericles A. Mitkas, ***"Representative Mining: The Synergy of Representative and DataMining"***, IEEE Intelligent Systems 2012.
4. Stuart Russell and Peter Norvig ***"Artificial Intelligence: A Modern Approach"*** 2011 Prentice-Hall, Inc.

5. John F. Elder IV, Dean W. Abbott, ***"A Comparison of Leading Data Mining Tools"*** Fourth International Conference Knowledge Discovery and Data Mining" New York, USA.
6. Philip Matkovsky, Dean W. Abbott, John F. Elder, ***"An Evaluation of High-end Data Mining For Fraud Detection"*** 2012.
7. T. Pharmine the data mining ***"Data Mining Tool Comparison"*** - Summary"
8. Simmi Bagga, G.N. Singh ***"Comparison of Data Mining And Auditing Tools"*** International Journal of Computer Science and Communication.
9. Dr. T.R. Gopalakrishnan Nair, Lakshmi Madhuri, Sharon Christa, Dr. V. Suma, ***"Data Preprocessing Model Using Intelligent Representative"*** International Conference on Information Systems Design and Intelligent Applications-2012
10. Steven F. Railsback, Steven L. Lytinen, Stephen K. Jackson, ***"Representative-based Simulation Platforms: Review and Development Recommendations"*** In press at Simulation. http://www-cdr.stanford.edu/ProcessLink/papers/JATL.html