Data Mining and MultiAgent Integration

Dr.NAGARATNA P.HEGDE
Professor
Vasavi Engineering College, hyd

B.VARIJA
Associate Professor
Nishitha college of eng and technology,RR

Abstract: *Data Mining and MultiAgent Integration attempts to present the latest attempts and trends in agent mining, rather than to cover the field in a dogmatic manner. To this, it has been divided into three parts. It provides an overview on the integration of agents and data mining, giving an inside view on the expected benefits and practical problems addressed upon integration, and presents a number of representative data mining-driven agents, carefully selected in order to cover a wide scope of applications and domains. Finally, focuses on challenges on MULTI Agent-driven data mining,*

Keywords:Data mining, multi agent system, agents,agent mining

1. INTRODUCTION

We are living in the data age. With huge amounts of data being produced around the world every day, here comes the need to deal with it efficiently in order to extract useful information from it. Since the data comes from a diverse range of sources, including social networking sites, supply chains and government databases; it is usually unstructured following no particular format or layout. Hence, we need to process the incoming data to find useful information in it. Data mining is the process used where in intelligent methods to extract interesting data patterns and knowledge from large amounts of data . However, the rate at which data is produced is very high, and we need efficient methods of mining to keep abreast with it. To ensure higher performance, we use the concept of agents to support the data mining process known as agent mining. The distributed nature of agent mining brings several advantages to data mining such as autonomy, scalability, and reliability, and security, interactivity and high speed . Agents can be used to automate the various tasks like data selection, data cleansing, and data pre-processing, to perform classification, clustering and knowledge representation. As an emerging area, a lot of research can be performed in this field. The main areas of research include agent-based data warehouse, agents for information retrieval, agents for distributed and parallel learning, information gathering agents and mobile agents for distributed data mining. In this paper, it describes the concept of an agent, explains agent mining. Along with the multi- agent data mining. this paper describes the concepts of multi-agent systems describes the existing multi-agent data mining systems, each benefitting the data mining process in its own way.

2. DATA MINING

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.

Most companies already collect and refine massive quantities of data. Data mining techniques can be implemented rapidly on existing software and hardware platforms to enhance the value of existing information resources, and can be integrated with new products and systems as they are brought on-line.

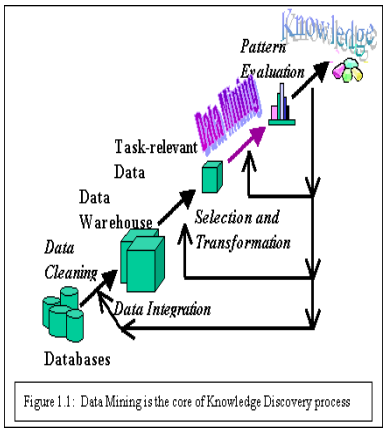


Figure 1.1: Data Mining is the core of Knowledge Discovery process

The Foundations of Data Mining

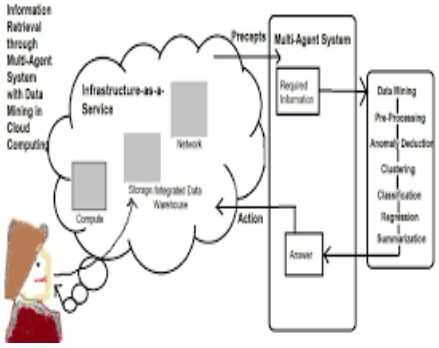
Data mining techniques are the result of a long process of research and product development. This evolution began when business data was first stored on computers, continued with improvements in data access, and more recently, generated technologies that allow users to navigate through their data in real time. Data mining takes this evolutionary process beyond retrospective data access and navigation to prospective and proactive information delivery. Data mining is ready for application in the business community because it is supported by three technologies that are now sufficiently mature:

- Massive data collection
- Powerful multiprocessor computers
- Data mining algorithms

Commercial databases are growing at unprecedented rates. A recent META Group survey of data warehouse projects found that 19% of respondents are beyond the 50 gigabyte level, while 59% expect to be there by second quarter of 1996.1 in some industries, such as retail, these numbers can be much larger. The accompanying need for improved computational engines can now be met in a cost-effective manner with parallel multiprocessor computer technology. Data mining algorithms embody techniques that have existed for at least 10 years, but have only recently been implemented as mature, reliable, understandable tools that consistently outperform older statistical methods.

In the evolution from business data to business information, each new step has built upon the previous one. For example, dynamic data access is critical for drill-through in data navigation applications, and the ability to store large databases is critical to data mining. From the user's point of view, the four steps listed in Table 1 were revolutionary because they allowed new business questions to be answered accurately and quickly

3. MULTI-AGENT DATA MINING SYSTEMS



Pikater Pikater is a multi-agent system that makes use of metalearning to perform a data mining task. The Pikater system enables a user to gain knowledge from a never-before-seen dataset by suggesting the best possible data mining method for mining this new dataset, obtained by meta learning over the previously obtained task results. The system stores data as well as metadata, which sometimes needs to be explicitly specified by the user with the task to be performed. This metadata consists includes the number of attributes in the datasets, number of records in the datasets, data type of the attributes and the missing values. Based on the data and metadata, the closest dataset to the dataset provided by the user is determined and the method used to mine that dataset is selected.

The Pikater multi-agent system is being developed using JADE framework. The system consists of four layers - user interface layer, computational layer, data layer and administrative layer.The task to be performed is defined by the user in a human understandable language to the UI layer. This layer is managed by the UI agents, which translates the task defined in a human understandable language into ontologies, and communicates it to the system. The computational layer consists of the data mining methods and is managed by the computing agent. The reader agents, used to read data from files and data manager service, used to access data stored in the database, make up the data layer. The agent and agent option managers control the administrative layer, which controls the entire problem solving process.

The agent manager acts as a link between the interface layer and computational layer, chooses the best method from the existing mining methods as well

as collects the results and provides statistical information. Four layer abstract architecture and the agents that constitute that layer Apart from providing a convenient, user-friendly and reusable system for data mining, the system suggests a data mining method for datasets that have not been encountered before. The use of multi-agent technology provides extensibility by allowing addition of new components to the existing system by use of structured ontology language and the international standards of agent's communication

Data Mining is the integration of multi-agent systems and distributed data mining wherein the concept of cooperative agents is used in data mining to overcome the challenges faced in a distributed environment like limited bandwidth, sensitivity of confidential data, limited distributed computing resources and complexity concerning multiple large systems generating huge amounts of data.

4. TYPES OF AGENTS

Interface agent: The interface agent is responsible for communication with the user, which includes accepting the task to be performed as input and providing the results as output. It is responsible for inter-agent communication.

Agent manager: On receiving a request from the interface agent, the agent manager forms a plan to complete the request. The agent manager is responsible for the completion of the user request, which it attains by assigning the work to different agents. The results are communicated to the interface agent. It is responsible for synchronization of the agents.

Data agent: The main function of the data agent is to supply data from multiple sources to the mining agent. The data agent maintains the metadata information of all the data sources.

Mining agent: The mining agent implements the mining algorithm. The mining agent initiates the mining technique based on the information provided by the knowledge module, such as the appropriate type of method for the problem at hand, the requirements of the method, form of input data, etc.

Result agent: The result agent receives the data mining result from the mining agent. The result agent is responsible for the presentation and visual representation of the knowledge with the help of the visualization primitives and report templates it maintains.

Broker agent: The broker agent contains the names, ontology and capabilities of all the agents registered with it. On receiving a request, the broker agent provides the corresponding names of the agents in order to fulfil the request.

Query agent: A query agent is created for each user request. The query agent uses the knowledge module schemas to generate queries in order to complete a user request.

6. AGENT MINING

The Synergy of Agents and Data Mining give an overall perspective of the driving forces, theoretical underpinnings, main research issues, and application domains of this field, while addressing the state-of-the-art of agent mining research and development. Our review is divided into three key research topics: agent-driven data mining, data mining-driven agents, and joint issues in the synergy of agents and data mining. This new and promising field exhibits a great potential for groundbreaking work from foundational, technological and practical perspectives.

A Bounded and Adaptive Memory-Based Approach to Mine Frequent Patterns from Very Large Databases could use only a bounded portion of the primary memory and this gives the opportunity to assign other parts of the main memory to other tasks with different priority. In other words, we propose a specialized memory management system which caters to the needs of the ARM model in such a way that the proposed data structure is constructed in the available allocated primary memory first. If at any point the structure grows out of the allocated memory quota, it is forced to be partially saved on secondary memory.

The secondary memory version of the structure is accessed in a block-by-block basis so that both the spatial and temporal localities of the I/O access are optimized. Thus, the proposed framework takes control of the virtual memory access and hence manages the required virtual memory in an optimal way to the best benefit of the mining process to be served. Several clever data structures are used to

facilitate these optimizations. All the above discussed methods have the problem of data retrieval and its accuracy about time complexity. Proposed Method The proposed multi agent system has three components namely Lookup manager, Agent Scheduler and GA Based Evolution. The lookup manager identifies set of locations where the data is

Basic HTML view

7. BENEFITS

Data mining and multi agent integration

- Integrates a multi-class supervised classification algorithm with agent technology in the domain of network intrusion detection, where a multi-agent design methodology is coupled with a highly accurate, fast, and lightweight PCC Classifier and CRSPM schemes.

- Proposes genetic algorithms for data extraction based on the evolution and grammatical composition of regular expressions
- Employs a weight-driven network module in order to increase projection of knowledge nodes in a system, enrich their repositories and stimulate the corresponding user communities.
- Defines the notion of Goal mining and utilizes it in order to extract knowledge on user goals, residing in common query logs.
- Finally, concludes the collection of manuscripts on data mining driven agents, discussing an agent-based diagnostic workbench equipped with classification capabilities, in order to support real medical diagnosis.

8. THE CHALLENGES OF DATA MINING-DRIVEN AGENTS

The astonishing rates at which data is generated and collected by current applications is difficult to handle even for the most powerful of today's computer systems. This windfall of information often requires another level of distillation to elicit the knowledge that is hidden in voluminous data repositories. Data mining can be used to extract knowledge nuggets that will constitute the building blocks of agent intelligence.

Here, intelligence is defined loosely so as to encompass a wide range of implementations from fully deterministic decision trees to self-organizing communities of autonomous agents. In many ways, intelligence manifests itself as effi- ciency. In rudimentary applications, agent intelligence is based on relatively simple rules, which can be easily deduced or induced, compensating for the higher development and maintenance costs.

In more elaborate environments, however, where both requirements and agent behaviors need constant modification in real time, these approaches prove insufficient, since they cannot accommodate the dynamic transfer of DM results into the agents. To enable the incorporation of dynamic, complex, and re-usable rules in multi-agent applications, a systematic approach must be adopted. Existing application data (i.e., past transactions, decisions, data logs, agent actions, etc.) can be filtered in an effort to refine the best, most successful, empirical rules and heuristics.

The resulting knowledge models can be embedded into 'dummy' agents in a process equivalent to agent training. As more data is gathered, the dual process of knowledge discovery and intelligence infusion can be repeated periodically, or on demand, to further improve agent reasoning. In data mining-driven agent systems, induction attempts to transform specific data and information into generalized knowledge models. During the induction process, new rules and correlations are produced, aimed at validating user hypotheses.

Since induction is based on progressive generalizations of specific examples, it may lead to invalid conclusions. In contrast, deductive systems draw conclusions by combining a number of premises. Under the assumption that these premises are true, deductive logic is truth preserving. In MAS applications, deduction is used when business rules and agent goals are well-defined and the human expert, who constructs the knowledge base, has a fine grasp of the problem's underlying principles. Nevertheless, deduction proves inefficient in complex and versatile environments. The coupling of the above two approaches usually leads to enhanced and more efficient reasoning systems.

Indeed, this combination overcomes the limitations of both paradigms by using deduction for well-known procedures and induction for discovering previously

unknown knowledge. The process of transferring DMextracted knowledge into newly-created agents is suitable for either upgrading an existing, non-agent-based application by adding agents to it, or for improving the already operating agents of an agent-based application. We consider three distinct cases, which correspond to three types of knowledge extracted and to different data sources and mining techniques. •

Case 1. Knowledge extracted by performing DM on historical datasets which record the business logic (at a macroscopic level) of a certain application;

• Case 2. Knowledge extracted by performing DM on log files recording the behavior of the agents (at a microscopic level) in an agent-based application, and

• Case 3. Knowledge extracted by the use of evolutionary data DM techniques in agent communities. 24 Longbing Cao In each case the software methodology must ensure:

a) the ability to dynamically embed the extracted knowledge models into the agents, and

b) the ability to repeat the above process as many times as is deemed necessary. Standard agent-oriented software engineering processes are followed, in order to specify the application ontology, the agent behaviors and agent types, the communication protocol between the agents, and their interactions.

A number of agent-based applications that cover all three cases of knowledge diffusion have been developed. Domains, that are better suited for

Case 1, include the traditional data producers, such as enterprise resource planning and supply chain management systems, environmental monitoring through sensor networks, and security and surveillance systems.

A typical example of Case 2 knowledge diffusion involves the improvement of the efficiency of agents participating in e-auctions. The goal here is to create both rational and efficient agent behaviors, which, in turn, will enable reliable agentmediated transactions. Another example is a web navigation engine, which tracks user actions in corporate sites and suggests possibly interesting sites. This framework can be extended to cover a large variety of web services and/or intranet applications.

Finally, Case 3 encompasses solutions for ecosystem modeling and for web

The need to study common issues for agent mining A typical issue is the involvement of human intelligence and human roles. Even though both communities recognize the importance of human involvement and human intelligence in problem-solving and solution development, it is challenging to effectively and dynamically include human roles in problem-solving systems. Issues arise from aspects such as the understanding and simulation of human empirical intelligence and experiences that are of critical importance to problem-solving, acquisition and representation of human qualitative intelligence in agent-mining systems, and the interaction and interfaces between humans and systems to cater for human intelligence and roles. Organizational, environmental and social factors constitute important elements of complex problems/systems and their environments in agents and data mining fields. This consists of comprehensive factors such as business processes, work-flows, business rules and human roles that are relevant to business problem-solving, organizational and social factors such as organizational rules, protocols and norms. For instance, while concepts such as organizational rules, protocols and norms have been fed into agent organizations, they are also important for data mining systems in converting patterns into operable deliverables that can be smoothly taken over by business people and integrated into business systems.

There are often gaps between technical outcomes and business expectations in developing workable agents and data mining algorithms and systems, which is due to the inconsistency and incompleteness of evaluation systems between technical and business concerns. As a result, the resulting deliverables are often not of interest to business people and are not operable for action-taking in business

problemsolving. An ideal scenario is to generate algorithms and systems that care about concerns from both the technical and business aspects and from both objective and subjective perspectives. The above case studies show that it is essential to study common issues for the benefits of the particular field. In fact, the studies can also activate the possible emergence of agent-mining symbionts.

For instance, the modeling and representation of domain knowledge and knowledge management in agents and data mining may be shared. It may serve as an intrinsic working mechanism for an agent-mining symbiont that has the capability of involving domain knowledge in agent-human interaction and data mining algorithm modeling, and managing knowledge for data mining agents and agent-based systems. To facilitate the studies of common enhancement issues, the possible methodologies and approaches needed may be highly diversified and cross-disciplinary.

9. CONCLUSION

This paper thus, gives a brief idea about the concept of agents and the multi-agent mining systems. Data mining in a distributed heterogeneous environment becomes flexible, adaptable, robust and easier with the use of data mining agents. Using the multi-agent systems for the data mining process we can tackle a large amount of information and also increase the speed of dealing with that information. we state that the concept of agent mining and the multi-agent systems have gained a huge momentum in the recent years and have a capability of delivering far more. More research in this respect can develop the data mining systems to a greater extent making them more efficient and increasing the accuracy of the mined data.

REFERENCES:

[1] G. Piatetsky-Shapiro, "Data Mining and Knowledge Discovery 1996 to 2005", 15(1), 99-105.

[2] P. Cabena, P. Hadjinian, R. Stadler, J. Verhees and A. Zanasi, "Discovering Data Mining: From Concept to Implementation", Prentice Hall, Upper Saddle River, Nj, 1998.

[3] D. Hand, H. Mannila and P. Smyth, "Principles of Data Mining", Mit Press, Cambridge, Ma, 2001.

[4] S. P. Deshpande1 And V. M. Thakare International Journal of Distributed and Parallel Systems (IJDPS) Vol.1, No.1, September 2010

[5] M. J. Berry, G. S. Linoff, "Data Mining Techniques for Marketing, Sales, and Customer Relationship, Management", Second Edition, 2004.

[6] O. R. Zaïane, "Principles of Knowledge Discovery in Databases", Cmput690, University Of Alberta, 1999.

[7] Ph. Preux, « Fouille De Données : Notes De Cours », Université De Lille 3, 9 Octobre 2008.

[8] G. Dong, J. Pei, 'Sequence Data Mining', Springer Edition, 2007.

[9] S. Prabhu, N. Venkatesan, "Data Mining And Warehousing", New Age International (P) Ltd., Publishers, New Delhi, 2007.

[10]J. Dasilva, C. Giannella, R. Bhargava, H. Kargupta, and M. Klusch. "Distributed Data Mining and Agents. Engineering Applications of Artificial Intelligence",