RETRO-DYNAMICS AND E-BUSINESS MODEL APPLICATION FOR DISTRIBUTED DATA MINING USING MOBILE AGENTS

Ezendu Ifeanyi Ariwa
Department of Accounting, Banking & Financial Systems
London Metropolitan University, United Kingdom

Mohamed B. Senousy
Associate Professor
Department of Computer and Information Systems
Sadat Academy for Management Sciences, Egypt

Mohamed M. Medhat
Department of Computer and Information Systems
Sadat Academy for Management Sciences, Egypt

Keywords: Knowledge Discovery – OIKI DDM – Decision Support System [DSS]

Abstract: Distributed data mining (DDM) is the semi-automatic pattern extraction of distributed data sources. The next generation of the data mining studies will be distributed data mining for many reasons. First of all, most of the current used data mining techniques require all data to be resident in memory, i.e., the mining process must be done at the data source site. This is not feasible for the exponential growth of the data stored in organization(s) databases. Another important reason is that data is inherently distributed for fault tolerance purposes. DDM requires two main decisions about the DDM implementations: A distributed computation paradigm (message passing, RPC, mobile agents), and the used integration techniques (Knowledge probing, CDM) in order to aggregate and integrate the results of the various distributed data miners. Recently, the new distributed computation paradigm, which has been evolved as mobile agent is widely used. Mobile agent is a thread of control that can trigger the transfer of arbitrary code to a remote computer. Mobile agents paradigm has several advantages: Conserving bandwidth and reducing latencies. Also, complex, efficient and robust behaviours can be realized with surprisingly little code. Mobile agents can be used to support weak clients, allow robust remote interaction, and provide scalability. In this paper, we propose a new model that can benefit from the mobile agent paradigm to build an efficient DDM model. Since the size of the data to be migrated in the DDM process is huge, our model will overcome the communication bottleneck by using mobile agents paradigm. Our model divides the DDM process into several stages that can be done in parallel on different data sources: Preparation stage, data mining stage and knowledge integration stage. We also include a special section on how current e-business models can use our model to reinforce the decision support in the organization. A cost analysis in terms of time consumed by each minor process (communication or processing) is given to illustrate the overheads of this model and the other models.

1 INTRODUCTION

Since distributed data mining is an emerging field of study, the modeling of distributed data mining systems is one of the key research areas in the data mining studies. We propose a new DDM model called OIKI (Optimized Incremental Knowledge Integration) system.

In this paper, we will discuss a theoretical background of the distributed data mining motivation and definitions in Section (2). In Section (3), related work to our research is discussed. DDM models are studied in Section (4). Section (5)

Ifeanyi Ariwa E., B. Senousy M. and M. Medhat M. (2004).
RETRO-DYNAMICS AND E-BUSINESS MODEL APPLICATION FOR DISTRIBUTED DATA MINING USING MOBILE AGENTS.
In *Proceedings of the Sixth International Conference on Enterprise Information Systems*, pages 500-507
DOI: 10.5220/0002596405000507

discusses the notation and different cost functions for DDM models. Our model will be presented and studied in details in Section (6). The most popular e-business model and how it can benefit from our proposed model are discussed in Section (7). An analytical study for the proposed cost functions for each model is discussed in Section (8) Finally, conclusions and future work are presented in Section (9).

2 BACKGROUND

The explosive growth in data stored in databases and data warehouses has generated an urgent need for new techniques that can intelligently transform this huge amount of data into useful knowledge. Consequently, data mining has become an important research area (Chen et al.: 1996).

Data mining differs from other data analysis techniques in that the system takes the initiative to generate patterns by itself (Information Discovery Inc.: 1997). Therefore, it is an exploratory analysis system (Turkey: 1973).

Data mining is concerned with the algorithmic means by which patterns, changes, anomalies, rules and statistically significant structures and events in data are extracted from large data sets (Fayyad et al: 1998 and Grossman: 1999).

Data mining studies can be classified into two generations. Studies in the first generation have focused on which kinds of patterns to mine. Studies in the second generation have focused on how mining can interact with other components in the framework like DBMS (Johnson: 2000).

Two requirements dictate the need for distributed data mining: data may be inherently distributed for a variety of practical reasons including security and fault tolerant distribution of data and services or mobile platform. Also, the cost of transporting data to a single site is usually high and sometimes unacceptable (Prodromidis: 1999; Kargupta et al.: 2000). The second requirement is that many of the mining algorithms require all data to be resident in memory. This might be unfeasible for large data sets, because these learning algorithms do not have the capability to process this huge amount of data. Data partitioning is one of the popular solutions for this problem (Provost: 1997). Consequently, data in this case is artificially distributed (Malhi: 1998).

DDM offers techniques to discover knowledge in distributed data through distributed data analysis using minimal communication of data (Kargupta et al.: 2000). Typical DDM algorithms involve local data analysis from which a global knowledge can be extracted using knowledge integration techniques (Kargupta et al.: 2000).

3 RELATED WORK

In Davies et al. (1996), one of the recent works in DDM studies is concerned with an agent-based approach to data mining.

Several DDM systems have been proposed. In Kargupta et al. (1997), PADMA system has been presented. In Botia et al. (1998), the basic design and implementation guidelines in a generic data mining system have been studied. In Martin et al.(1999), an agent infrastructure for data mining systems has been proposed. In Stolfo (1997), Java Agents for Meta-learning (JAM) over distributed databases has been proposed. In Kargupta et al. (1999), Collective Data Mining (CDM) theory and implementation have been studied. In Chattratichat et al. (1999), architecture for distributed enterprise data mining has been presented.

In Guo (1999), a knowledge integration technique using knowledge probing has been studied.

The cost models for Client/Server, mobile agents and hybrid DDM models have been proposed in Krishnaswany et al. (2000). This study has presented a different data mining scenarios involving various architectural models.

4 DDM MODELS

There are two architectural models used in the development of DDM systems: Client/Server (CS) model and mobile agent model. In the following subsections, we will discuss each model.

4.1 Client/Server Based DDM Model

The Client/Server model uses the remote procedure call (RPC) mechanism in the communication between the clients and the server. The RPC allows a program on the client to invoke a procedure on the server using stubs on each side. The client-side stub acts as a proxy for the real procedure. It accepts calls for the procedure and arranges for them to be forwarded to the server. The server-side stub receives the call for a procedure and returns the results to the client-side stub. Finally, the client-side stub returns the result to the original RPC call (Crowley: 1997; Gray: 1995).

The CS-based DDM uses one or more DM servers. The client requests are sent to DM server that determines the required data sources and collects data from different locations and brings all the required data for the specified mining process to the DM sever. The DM server in turn houses the data mining algorithms. The mining process is accomplished on the DM server and the results are returned to the requested client.

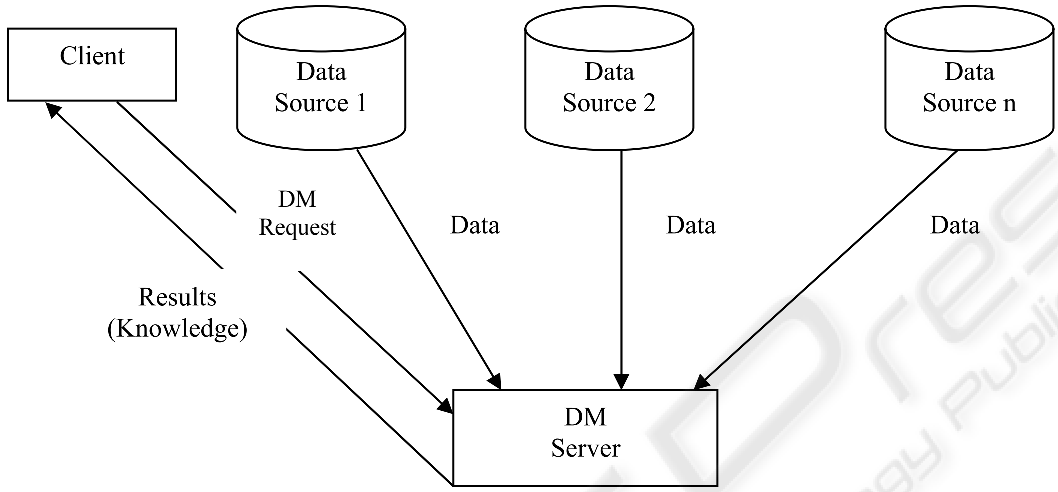

Figure 1: Illustrates typical Client/Server based DDM process

4.2 Mobile Agent Based DDM Model

A fundamental problem exists with Client/Server architectures is that if the server does not provide the exact service that the client requires, then the client must take a series of RPCs to obtain the end service. This might result in an overall latency increase and in intermediate information between the client and the server during the service processing. Consequently, CS architecture may waste the network bandwidth. Dale (1997) Gray et al.(2000). A mobile agent does not waste the bandwidth, because the agent migrates to the server. The agent performs the necessary sequence of operations locally, and returns just the final result to the client. Gray et al. (2000). The major drawback in the CS-based DDM model is that huge amount of data sets migrate form the data sources locations to the DM sever to accomplish the required DM process. This results into a considerable waste in the network bandwidth and consequently a big increase in latency.

A typical mobile agent-based DDM process begins with a client request for a DM process. The client determines the required data severs for the DM process and multicasts a set of mobile agents data miners MADMs. The MADMs migrate to the data servers and perform the data mining operations locally and return the final results (knowledge) to the client. Finally, the client uses a knowledge integration (KI) program to integrate the DM results from the different MADMs. Figure (2) illustrates the described mobile agent-based DDM process.

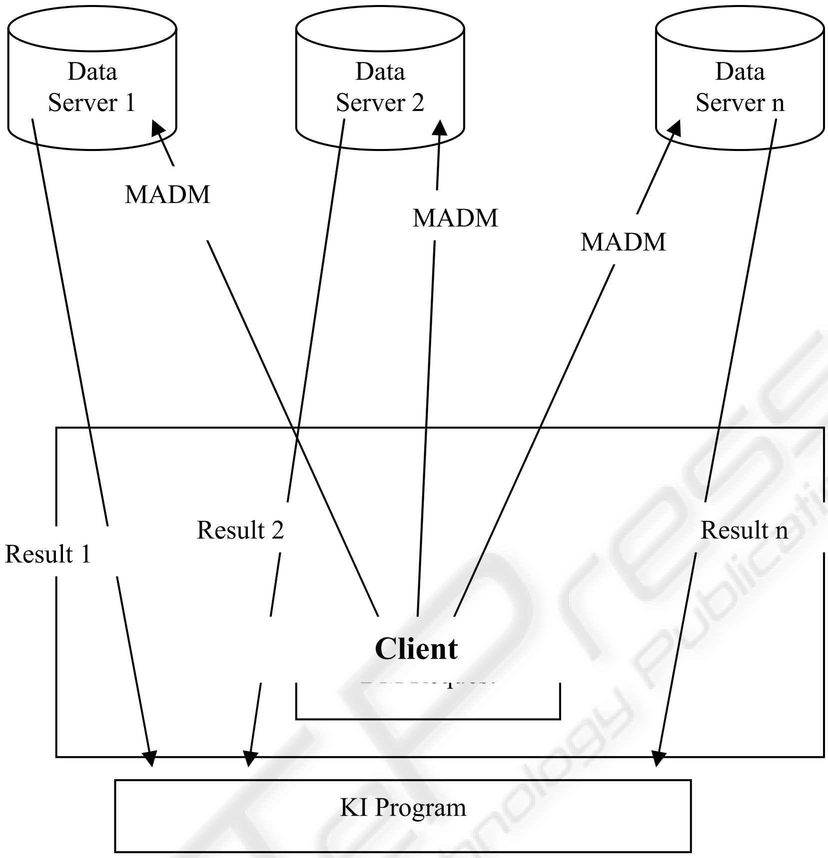

Figure 2: Typical mobile-agent-based DDM Process

5 OIKI DDM MODEL

Optimized Incremental Knowledge Integration (OIKI) DDM model is a mobile agent based DDM model that overcome the drawbacks of the traditional mobile agent based DDM model. Instead of transferring the results from each data server to the client, the client controls migration of the results among data servers to be integrated locally and finally, the final results are transferred to the client.

The typical OIKI DDM process: the client multicasts MADMs and MAKIs (Mobile Agents-Knowledge Integrators) to the required data servers. The data mining process is performed locally on each data server. The size of results of the first two accomplished DM processes are compared. The smaller one is migrated to the larger one. The knowledge integrator agent integrates the results of these two data servers. This process is repeated until all integrated results are resident in a specific data server and finally, the final results are sent back to the client. Consequently, the OIKI DDM process passes through three main stages: 1) Preparation Stage: The client multicasts MADMs and MAKIs to data servers. 2) Data Mining Stage: Data mining process is performed locally on each data server. 3) Knowledge Integration Stage: An incremental knowledge integration technique is performed on the data servers where the smaller results are migrated to the larger one to optimize the cost of results migration among data servers. Figure (3) illustrates the typical OIKI DDM process.

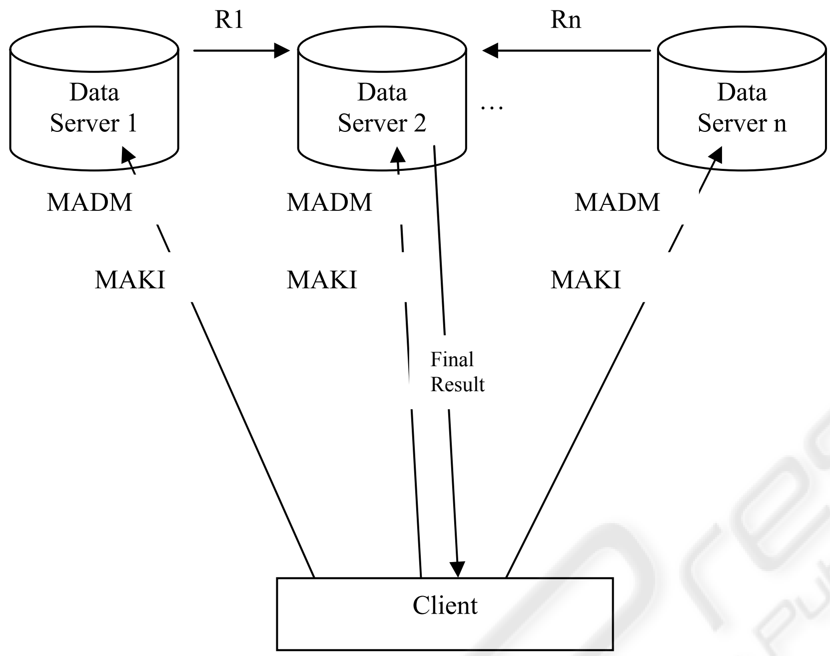

Figure 3: Typical OIKI DDM Process

6 THE APPLICATION OF OIKI DDM MODEL IN E-BUSINESS

The OIKI DDM model has its great benefits to current e-business models for many reasons; we explore here some facts to demonstrate our idea:

- Current e-business models depend on the existence of one or more databases in order to store the business data such as storefront model (Amazon.com), portal model (Yahoo.com), and recruiting on the web model (Monster.com). Due to the growth amounts of data stored in these databases, many databases are partitioned and distributed among several sites. Other organizations build a number of data marts to be used in the decision making process.
- Data mining could be used in Customer Relationship Management (CRM) as follows:
  a) Association rules extraction for customer attraction.
  b) Sequential patterns extraction for customer retention.
  c) Deriving classifiers and data clusters for cross-selling.

From the above discussion, the OIKI DDM model could be used to mine the business data efficiently. The MADMs and MAKIs are sent to the database partitions or data marts of a specific e-business organization and perform the data mining and knowledge integration processes locally. So, this model makes the data mining process scalable to any number of data sites. In addition, this model can be used to mine the data of several e-business organizations in the same field. These organizations have a contract to benefit from the hidden knowledge of their stored data. This can raise the efficiency of these organizations, because the extracted knowledge will be more accurate. The following are some examples of a typical use of OIKI DDM model in e-business:

Example 1

75% of customers purchase product 1 also purchase product 2 of an e-business organization.

This rule can be obtained using OIKI DDM model as follows:

- A client sends a set of mobile agents association rule mining (MADMs) and mobile agents association rule integrator (MAKIs) to the database partitions containing the needed data about the purchased products.
- MADMs perform the association rule mining process locally.

- MAKIs migrate with the results from a database partition containing smaller results to a database partition containing larger ones in order to optimize the communication cost.
- The previous step is repeated until all the results are integrated.

We should note that OIKI DDM model has a great advantage over the traditional mobile agent based models. The incremental integration of the mining results makes the strength of a rule increased or decreased incrementally. Thus, some rules may be disappeared before the knowledge integration process is finished.

Example 2

Clients purchasing product x, tend to be engineers.

This derived classifier could be obtained using OIKI DDM model:

- A client sends a set of mobile agents classification mining (MADMs) and mobile agents classifier integrator (MAKIs) to the database partitions containing the needed data about the purchased products and their customers.
- MADMs derives a set of classifier functions using each database partition as a training set.
- MAKIs migrate with classifiers from a database partition containing smaller classifiers to a database partition containing larger ones in order to optimize the communication cost. Meta-classification techniques might be used.
- The previous step is repeated until all the results are integrated.

The advantage of OIKI DDM model over the other DDM models is the use of a database partition as a training set which makes the time needed for training decreased. Then, the meta-classification process is done incrementally which makes this process easier and time efficient.

From the above example, the extracted knowledge would be used in marketing, so,

a) The organization places the product 2 to advertisements in the web page of product 1.

b) The organization sends newsletters and special offers about product x to engineers.

We can conclude that OIKI DDM model is used in various data mining techniques when data is distributed among several sites and it is more efficient over the other models because of the use of mobile agent technology and the incremental knowledge integration.

7 BENEFITS TO E-BUSINESS

There are a number of e-business models currently used in the implementation of e-business applications. Examples of such models are storefront model, auction model, portal model, dynamic pricing models, B2B models, online trading and lending models and e-learning models. The most commonly used model is storefront model using shopping-cart technology, where there is a merchant database stores all information about customers and goods. Deittel et al. (2001). And the e-business intelligence is accomplished through the use of data miners. Senousy et al. (2001).

The OIKI DDM model can be used in the storefront e-business model where the data servers are the merchant's databases. Thus, the merchant can analyze the data stored in the distributed databases used in the e-business. Figure (5) shows how the storefront e-business model can benefit from the OIKI DDM model in the data analysis. The e-business analyzer software sends MADMs and MAKIs to the required databases according to the DM request. The MADMs perform the data mining tasks on each database. The MAKIs perform the knowledge integration tasks such that the smaller results migrate to the larger ones. The e-business analyzer software controls the results migration among merchant's databases in order to optimize the incremental knowledge integration process.

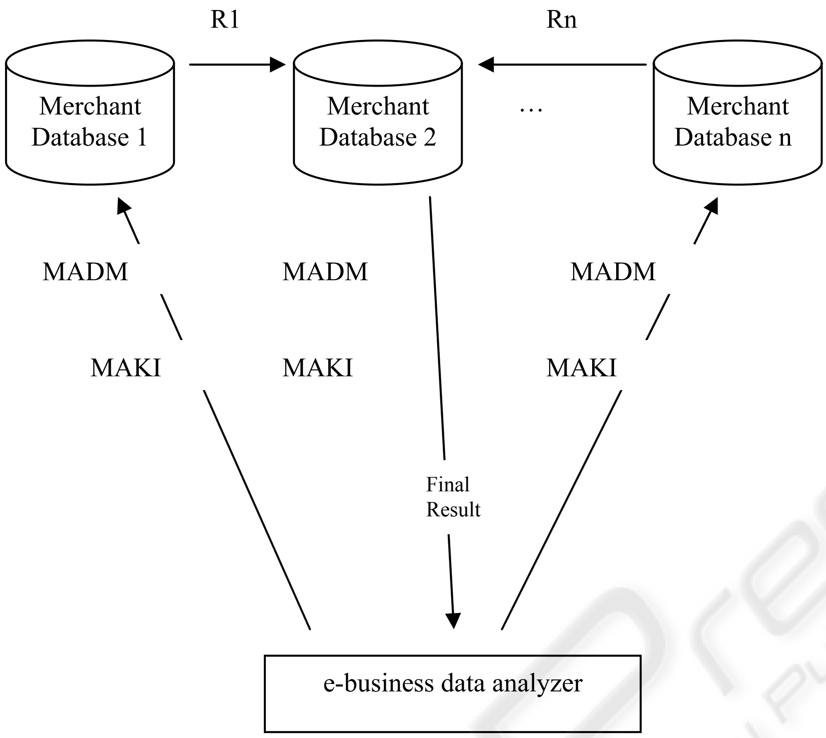

Figure 5: Applying OIKI DDM to storefront e-business model

8 CONCLUSIONS AND FUTURE WORK

The OIKI DDM model overcomes the drawbacks of traditional mobile agent based DDM model by making the knowledge integration process an incremental process. This makes the DDM system scalable to any number of data servers. The way of performing the incremental knowledge integration process in the OIKI DDM model makes this process optimized. The storefront e-business model can benefit from our model by applying the DDM process on the merchant's databases.

Our future work is concerned with an analytical based comparison among DDM models, and performance evaluation of these models. Advanced simulation techniques would be used. The implementation issues of the proposed model are research areas that have a lot of work to be done. The study of the efficient knowledge integration techniques is an essential research area to the OIKI DDM model implementation.

REFERENCES

Botia, A., Garijo, R., and Skarmeta F., 1998, "A Generic Data Mining System: Basic Design and Implementation Guidelines", Workshop on Distributed Data Mining at the 4th International Conference on Data Mining and Knowledge Discovery (KDD-98).

Chen, M., Han, J., and Yu, P., 1996, "Data Mining: An Overview from Database Perspective", IEEE Transactions on Knowledge and Data Engineering, 8(6): 866-883, 1996.

Crowley, C., 1997, Operating Systems: A Design-Oriented Approach, IRWIN, Boston.

Dale, J., 1997, "PhD thesis: A Mobile Agent Architecture to Support Distributed Resource Information Management", Department of Electronics and Computer Science, Faculty of Engineering, University of Southampton.

Davies, W., and Edwards, P., 1996, "Distributed Learning: An Agent-Based Approach to Data Mining", Technical Report of Department of Computer Science, King's College, University of Aberdeen.

Deitel, H., Deitel, P., and Neito, T., 2001, e-Business and e-Commerce: How to Program, Prentice Hall.

Fayyad, U., Bradley, P., and Mangasarian O., 1998, " Mathematical Programming for Data Mining: Formulations and Challenges", Journal of Computing, special issue on Data Mining.

Gray, R., Kotz, D., Cybenko, G., Rus, D., 2000, "Mobile agents: Motivations and state-of-the-art systems", ftp://ftp.cs.dartmouth.edu/TR/TR2000-365.ps.Z.

Gray, R., 1995, "Ph.D. Thesis Proposal: Transportable Agents", Department of Computer Science, Dartmouth College.

Guo, Y., and Sutiwaraphun, 1999, "Integrating Knowledge in Distributed Data Mining", Department of Computing, Imperial College.

Grossman, R., Kasif, S., Moore, R., Rocke, and Ullman, J., 1999, "Data Mining Research: Opportunities and Challenges", A Report of three Workshops on Mining Large, Massive, and Distributed Data.

Information Discovery Inc., 1997, "A Characterization of Data Mining Technologies and Processes", Journal of Data Warehousing.

Johnson, T., Lakshmanan, L., and Ng, R., 2000, "The 3W Model and Algebra for Unified Data Mining", Proceedings of the 26th VLDB Conference.

Kargupta, H., Hamzaoglu, I. and Stafford, B., 1997, "Scalable, Distributed Data Mining Using An Agent Based Architecture", in Proc. of the 3rd Int. Conf. on Knowledge Discovery and Data Mining, Newport Beach, California, (eds), D.Heckerman, H.Mannila, D.Pregibon, and R.Uthurusamy,

Kargupta,H., Park,B., Hershberger,D., and Johnson, E., 1999, "Collective Data Mining: A New Perspective Toward Distributed Data Mining", to appear in Advances in Distributed Data Mining, (eds) H.Kargupta and P.Chan, AAAI Press.

Krishnaswamy, S., Zaslavsky,A., and Loke,S,W., 2000, "An Architecture to Support Distributed Data Mining Services in E-Commerce Environments", 2nd International Workshop on Advanced Issues in E-Commerce and Web-Based Information Systems, San Jose, Californinia, July 8-9.

Malhi, B., 1998, "Master thesis report: Providing Support for Resource Management Tools in a Wide Area High Performance Distributed Data Mining System", Laboratory for Advanced Computing, University of Illinois at Chicago, http://lac.uic.edu/~balinder/thesis.htm.

Martin,G., Unruh,A., and Urban,S., 1999, "An Agent Infrastructure for Knowledge Discovery and Event Detection", Technical Report MCC-INSL-003-99, Microelectronics and Computer Technology Corporation (MCC).

Prodromidis, A., 1999, "Ph.D. Thesis: Management of Intelligent Learning Agents in Distributed Data Mining Systems", School of Arts and Science, Columbia University.

Provost, F., 1997, "Scaling Up Inductive Algorithms: An Overview", Proceedings of the Third International Conference on Knowledge Discovery and Data Mining, California, August, 1997, pp 239-242.

Senousy, M., and Medhat, M., 2001, "A Proposed Architecture for E-telligence Integration Model", Proceedings of the 8th Scientific Conference on Information Systems and Computer Technology, Cairo.

Stolfo,S,J., Prodromidis,A,L., Tselepis, L., Lee,W., Fan,D., and Chan,P,K., 1997, "JAM: Java Agents for Meta-Learning over Distributed Databases", in Proc. of the 3rd Int. Conf. On Data Mining and Knowledge Discovery (KDD-97), Newport Beach, California, (eds) D.Heckerman, H.Mannila, D.Pregibon, and R.Uthurusamy, AAAI Press, pp. 74-81.

Turkey, J., 1973, Exploratory Data Analysis, New York: McMillan.