Interaction and Integration of Agent Mining in Distributed Data Environment

K.Raghava Rao[1], M. Nagabhushnam Rao[2], Y.Siva Prasad [3] & K.Ruth Ramya

Professor & Head[1], MCA Department, Asst. Prof[3], CSE Department, Asst. Professor[4], Dept. of CSE,

KLUniversity,Vaddeswaram,AP State ,Professor & Head[2], BV Raju Engg. Colleg Bhimavaram,AP

{raghavarao1@yahoo.com},{mnraosir,sivaprasady,ruthmoses.mathi}@gmail.com

Abstract: In recent years, more and more researchers have been involved in research on both agent technology and distributed data mining. A clear disciplinary effort has been activated toward removing the boundary between them,that is the interaction and integration between agent technology and distributed data mining. We refer this to *agent mining* as a new area. The marriage of agents and distributed data mining is driven by challenges faced by both communities, and the need of developing more advanced intelligence, information processing and systems. In this paper presents an overall picture of agent mining from the perspective of positioning it as an emerging area. We summarize the main distributed data mining, driving forces, disciplinary framework, applications, and trends and directions, data mining-driven agents, and mutual issues in agent mining. Arguably, we draw the following conclusions: (1) agent mining emerges as a new area in the scientific family, (2) both agent technology and distributed data mining can greatly benefit from agent mining, (3) it is very promising to result in additional advancement in intelligent information processing and systems. However, as a new open area, there are many issues waiting for research and development from theoretical, technological and practical perspectives.

Keywords: Agent mining, distributed data mining& environment, KDD, AAMAS, DDM Algorithm, Adaptive Learning

1. Introduction

Autonomous agent and multi-agent systems (AAMAS, refer to here as *agents*) [44] and knowledge discovery from data (KDD, or otherwise known as *data mining*)[10] have emerged and developed separately in the last twenty years. Both areas are currently very active. Agents primarily focus on issues from many aspects, from theoretical, methodological, and experimental to practical issues in developing agent-based computing and agent-oriented intelligent systems, which are a powerful technology for autonomous intelligent system analysis, design and implementation. The major topics of interest consist of research on individual agents, multi-agent systems (MAS), methodology and techniques, tools and applications. The agent technology contributes to many diverse domains such as software engineering, user interfaces, ecommerce, information retrieval, robotics, computer games, education and training, ubiquitous computing, and social simulation.

Currently, agent studies have been spread from programming to organizational and societal factors to study agents and agent-based systems. The research on agents has far exceeded the original community scope of artificial intelligence and software. Researchers from many other areas have started to discuss, develop, wrap and use the concept of agents, covering almost all aspects of the social sciences such as law, business, organizational, behavior sciences, finance and economics, tourism, not to mention the extensive family of natural science and technology. The benefits from agents are expected to be very comprehensive and diverse, from academic disciplines, to the sciences, the social sciences and the humanities.

Similarly, *data mining* originally focused on knowledge discovery in databases, but it has experienced a migration from data-centered pattern discovery, to knowledge discovery, actionable knowledge discovery, and currently to domain-oriented decision delivery [11]. Data mining and its tools is becoming a ubiquitous information processing field and tools, involving techniques and researchers from many areas such as statistics, information retrieval, machine learning, artificial intelligence, pattern recognition, and database technologies. Data mining is increasingly widely tested in varying applications and domains, for instance, web mining and services, text mining, telecommunications, retail, governmental service, fraud, security, business intelligence studies. Besides the emphasis of in-depth data intelligence, recent efforts in data mining cover many additional areas and domain problems. Data mining researchers recognize the need to involve the environment, human intelligence, domain intelligence, organizational intelligence, and social intelligence in the mining process, models, the findings and deliverables.

This will trigger another wave of migration from the discovery of knowledge to the delivery of deep knowledge-based problem-solving systems and services. The above analysis of trends and directions of both areas shows that these two independent research streams have been created and originally evolved with separate aims and objectives. They used to target individual methodologies and techniques to cope with domain-specific problems and challenges in respective areas. However, both are concerned with many similar aspects and factors, such as human roles, user system interaction, dynamic modeling, domain factors, organizational and social factors. In fact, both areas contribute to the advancement of intelligence, and intelligent information processing, services and systems. In fact, they need each other, as evidenced by typical topics of agent-based data mining in the middle 1990s. Consequently, we see a clear trend of the interaction and integration between agents and data mining. Its development has reached the level of a new and promising area, and is moving towards becoming a first-class citizen in the science and technology family [12, 5, 6]. In this paper presents an overall picture of this emerging field, distributed data mining and multi-agent integration. We first analyze the respective and common challenges in agents and distributed data mining areas. These challenges motivate and drive the need and emergence of agent mining. A scientific framework and theoretical underpinnings are presented, which illustrate the synergy methods and foundations of agents and data mining. Further, we briefly summarize the research on three major directions in agent mining, namely agent-driven distributed data mining, mining-driven agents, and mutual issues in agent mining and applications are discussed. Finally, we discuss the development of agent mining community. Information provided here can benefit new researchers, and enable them to quickly step into this field.

2. Distributed Data Mining

Traditional warehouse-based architectures for data mining suppose to have centralized data repository. Such a centralized approach is fundamentally inappropriate for most of the distributed and ubiquitous data mining applications. In fact, the long response time, lack of proper use of distributed resource, and the Fundamental characteristic of centralized data mining algorithms do not work well in distributed environments. A scalable solution for distributed applications calls for distributed processing of data, controlled by the available resources and human factors. For example, let us suppose an ad hoc wireless sensor network where the different sensor nodes are monitoring some time-critical events. Central collection of data from every sensor node may create traffic over the limited bandwidth wireless channels and this may also drain a lot of power from the devices. A distributed architecture for data mining is likely aimed to reduce the communication load and also to reduce the battery power more evenly across the different nodes in the sensor network. One can easily imagine similar needs for distributed computation of data mining primitives in ad hoc wireless networks of mobile devices like PDAs, cell phones, and wearable computers. The wireless domain is not the only example. In fact, most of the applications that deal with time-critical distributed data are likely to benefit by paying careful attention to the distributed resources for computation, storage, and the cost of communication. As another example, let us consider the World Wide Web: it contains distributed data and computing resources. An increasing number of databases (e.g., weather databases, oceanographic data, etc.) and data streams (e.g., financial data, emerging disease information, etc.) are currently made on-line, and many of them change frequently. It is easy to think of many applications that require regular r monitoring of these diverse and distributed sources of data. A distributed approach to analyze this data is likely to be more scalable and practical particularly when the application involves a large number of data sites. Hence, in this case we need data mining architectures that pay careful attention to the distribution of data, computing and communication, in order to access and use them in a near optimal fashion. Distributed Data Mining (sometimes referred by the acronym DDM) considers data mining in this broader context. DDM may also be useful in environments with multiple compute nodes connected over high speed networks. Even if the data can be quickly centralized using the relatively fast network, proper balancing of computational load among a cluster of nodes may require a distributed approach. The privacy issue is playing an increasingly important role in the emerging data mining applications. For example, let us suppose a consortium of different banks collaborating for detecting frauds. If a centralized solution was adopted, all the data from every bank should be collected in a single location, to be processed by a data mining system. Nevertheless, in such a case a distributed data mining system should be the natural technological choice: both it is able to learn models from distributed data without exchanging the raw data between different repository, and it allows detection of fraud by preserving the privacy of every bank's customer transaction data. For what concerns techniques and architecture, it is worth noticing that many several other fields influence Distributed Data Mining systems concepts. First, many DDM systems adopt the Multi-Agent System (MAS) architecture, which finds its root in the Distributed Artificial Intelligence (DAI). Second, although Parallel Data Mining often assumes the presence of high speed network connections among the computing nodes, the development of DDM has also been influenced by the PDM literature. Most DDM algorithms are designed upon the potential parallelism they can apply over the given distributed data.. In figure 1 a general Distributed Data Mining framework is presented. In essence, the success of DDM algorithms lies in the aggregation. Each local model represents locally coherent patterns, but lacks details that may be required to induce globally meaningful knowledge. For this reason, many DDM algorithms require a centralization of a subset of local data to compensate it. The ensemble approach has been applied in various domains to increase the accuracy of the predictive model to be learnt. It produces multiple models and combines them to enhance accuracy. Typically, voting (weighted or un-weighted) schema are employed to aggregate base model for obtaining a global model. As we have discussed above, minimum data transfer is another key attribute of the successful DDM algorithm.

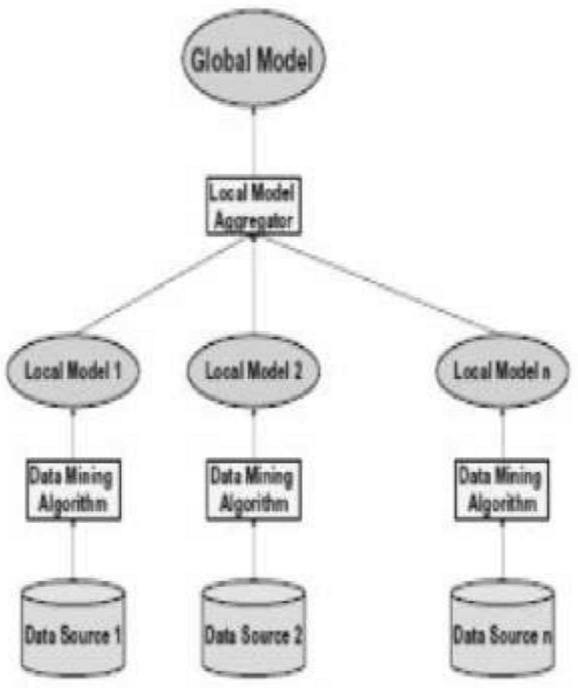


Figure 1: General Distributed data mining Frame work

3. The Challenges of Distributed Data Mining

Data mining and machine learning currently forms a mature field of artificial intelligence supported by many various approaches, algorithms and software tools. However, modern requirements in data mining and machine learning inspired by emerging applications and information technologies and the peculiarities of data sources are becoming increasingly tough. The critical features of data sources determining such requirements are as follows:

In enterprise applications, data is distributed over many heterogeneous sources coupling in either a tight or loose manner. Distributed data sources associated with a business line are often complex, for instance, some is of high frequency or density, mixing static and dynamic data, mixing multiple structures of data; Data integration and data matching are difficult to conduct; it is not possible to store them in centralized storage and it is not feasible to process them in a centralized manner; In some cases, multiple sources of data are stored in parallel storage systems; Local data sources can be of restricted availability due to privacy, their commercial value, etc., which in many cases also prevents its centralized processing, even in a collaborative mode; In many cases, distributed data spread across global storage systems is often associated with time difference; Availability of data sources in a mobile environment depends on time; The infrastructure and architecture weaknesses of existing distributed data mining systems requires more flexible, intelligent and scalable support. These and some other peculiarities require the development of new approaches and technologies of data mining to identify patterns in distributed data. Distributed data mining (DDM), in particular, Peer- to-Peer (P2P) data mining, and multi-agent technology are two responses to the above challenges.

4. Challenges in Distributed Data Mining Disciplines

Data mining faces many challenges when it is deployed to real world problem solving, in particular, in handling complex data and applications. We list here a few aspects that can be improved by agent technology. These include enterprise data mining infrastructure, involving domain and human intelligence, supporting parallel and distributed mining, data fusion and preparation, adaptive learning, and interactive mining.

(a) Enterprise data mining infrastructure: The development of data mining systems supporting real-world enterprise applications are challenging. The challenge may arise from many aspects, for instance, integrating or mining multiple data sources, accessing distributed applications, interacting with varying business users, and communicating with multiple applications. In particular, it has been a grand challenge and a longstanding issue to build up a distributed, flexible, adaptive and efficient platform supporting interactive mining in real-world data.

(b) Involving domain and human intelligence: Another grand challenge of existing data mining methodologies and techniques are the roles and involvement of domain intelligence and human intelligence in data mining. With respect to domain intelligence, how to involve, represent, link and confirm to components such as domain knowledge, prior knowledge, business process, and business logics in data mining systems is a research problem. Regarding human intelligence, we need to distinguish the role of humans in specific applications, and further build up system support to model human behavior, interact with humans, bridge the communication gap between data mining systems and humans, and most importantly incorporate human knowledge and supervision into the system.

(c) Data fusion and preparation: In the real world, data is getting more and more complex, in particular, sparse and heterogeneous data distributed in multiple places. To access and fuse such data needs intelligent techniques and methods. On the other hand, today's data preparation research is facing new challenges such as processing high frequency time series data stream, unbalanced data distribution, rare but significant evidence extraction from dispersed data sets, linking multiple data sources, accessing dynamic data. Such situations expect new data preparation techniques.

(d) Adaptive learning: In general, data mining algorithms are predefined to scan data sets. In real-world cases, it is expected that data mining models and algorithms can adapt to dynamic situations in changing data based on their self-learning and self-organizing capability. As a result, models and algorithms can automatically extract patterns in changing data. However, this is a very challenging area, since existing data mining methodologies and techniques are basically non-automatic and inadaptable. To enhance the automated and adaptive capability of data mining algorithms and methods, we need to search for support from external disciplines that are related to automate and adaptive intelligent techniques.

(e) Interactive mining: Controversies regarding either automatic or interactive data mining have been raised in the past. A clear trend for this problem is that interaction between humans and data mining systems plays an irreplaceable role in domain-driven data mining situations. In developing interactive mining, one should study issues such as user modeling, behavior simulation, situation analysis, user interface design, user knowledge management, algorithm/model input setting by users, mining process control and monitor, outcome refinement and tuning. However, many of these tasks cannot be handled by existing data mining approaches.

5. Driving Forces of Agent Mining Interaction and Integration

The emergence of agent mining results from the following driving forces: The critical challenges in agents and data mining respectively, the critical common challenges troubling agents and data mining the complementary essence of agents and data mining in dealing with their challenges, and the great add-on potential resulting from the interaction and integration of agents and data mining. Agents and data mining are facing critical challenges from respective areas. Many of these challenges can be tackled by involving advances in other areas.

In this section, we specify both individual and mutual challenges in agent and mining disciplines that may be complemented by the interaction with the other disciplines.

6. Challenges in Agent Disciplines

As addressed in some retrospective publications, traditional agent technology has been challenged in many aspects such as developing organizational and social intelligence.

Figure **2:** Challenges in agents and data mining.

In the following analysis, we explain this from the following aspects: agent awareness, agent learning, agent action ability, and agent distributed processing, agent in-depth services, and agent constraint processing.

Fig.2. illustrates these challenges.

(a) Agent awareness: Agent awareness refers to the capability of an agent to recognize internal and/or external environment change, and analyze situation change. In contrast to normal sensing and perception as conducted in reactive agents, here agent awareness specifically refers to situation analysis and environment modeling driven by agent learning and discovery. Agents with such a capability should self-recognize, compare and reason the changes taking place in the environment. To this end, it is necessary for agents to accumulate learning capability.

(b) Agent learning: In open multi-agent organizations, interaction widely exists between agent and environment, and between an agent and the other agent(s). Agents are expected to learn from other agents, their environment, and from the interaction and dynamics. In addition, agents may be expected to learn from users and interaction with humans. To foster such learning capability, agents need to be fed with learning and reasoning algorithms that can support them to discover reason or simulate interesting information from

interactive and situational data. Learning capability is widely recognized to be significant for enhancing agent intelligence. On the basis of the varying objectives, agent learning has been paid unprecedented attention. Multiple forms of agent learning capability are being studied. Agent learning may be conducted in terms of agent architectures such as cognitive learning, deductive learning, distributed learning, and cooperative learning. With respect to learning objectives, agent learning may also be classified into procedural learning, action learning, rule and pattern learning, and decision-making learning. From the learning process aspect, agent learning can be categorized into reinforcement learning, discovery learning, single-trial learning, reasoning learning, and random learning. The implementation of agent learning presents either a passive or active manner. In a broad sense, learning can be in a supervised, unsupervised or hybrid manner.

(c) Agent action ability: Agent action ability refers to the capability of an agent to take actions to its advantage on the basis of the knowledge obtained through in-depth analysis, reasoning and discovery. Unlike general action taken by agents, we are specifically interested in actions for recommendation, servicing, searching, discovery, conflict resolution, etc. with great benefits but low costs. To this end, agents need to balance benefits and costs, and maximize their return while minimizing the risk before taking an action or a sequence of actions.

(d) Agent distributed processing: In middle to large scale multi-agent systems, agents need to deal with distributed processing tasks such as learning from agents across multiple organizations, applications or data, conducting decentralized coordination, cooperation and negotiation among agents crossing resources, and implementing information gathering, dispatching and transport among agents located in distributed applications. To tackle the above tasks in distributed conditions, agents need to make decisions after analyzing and utilizing relevant information from multiple sources. Information analysis and utilization is not a trivial job. Agents may need to develop capabilities such as data analysis and discovery, procedural learning, goal adjustment, and information fusion.

(e) Agent in-depth services: Agents are often developed for providing varied services, for instance, network services such as web recommender systems, mobile agents for information searching and passing, and user services such as for user interaction and user modeling. Smart service providing relies on in-depth analysis of the service request-related data and information, as well as service historical data and service performance, in order to deeply understand service data and select the best service solutions. However, the agent community often does not work on such kinds of capabilities.

(f) Agent constraint processing: Open complex agent systems often involve many types of constraints from many aspects, for instance, temporal and spatial constraints, or execution constraints from organizational aspects. Such constraints form conditions in improving agent capabilities such as learning, adaptation, action ability, and services. There is a need to understand such constraints, and to involve and best treat such constraints in an agent system and solution generation.

7. Mutual Challenges in Agent and Distributed Data Mining

As addressed in [5, 6, 7], agents can enhance data mining through involving agent intelligence in data mining systems, while an agent system can benefit from data mining via extending agents' knowledge discovery capability. Nevertheless, the agent mining interaction symbiosis cannot be established if mutual issues are not solved. These mutual issues involve fundamental challenges hidden on both sides and particularly within the interaction and integration. Figure 2 presents a view of issues in agent-mining interaction highlighting the existence of mutual issues. Mutual issues constraining agent-mining interaction and integration consist of many aspects such as architecture and infrastructure, constraint and environment, domain intelligence, human intelligence, knowledge engineering and management, and nonfunctional requirements.

(a) Architecture and infrastructure: Data mining always faces a problem in how to implement a system that can support those brilliant functions and algorithms studied in academia. The design of the system architecture conducting enterprise mining applications and emerging research challenges needs to provide (1) functional support such as crossing source data management and preparation, interactive mining and the involvement of domain and human intelligence, distributed, parallel and adaptive learning, and plug-and-play of algorithms and system components, as well as (2) nonfunctional support for instance adaptability, being user and business friendly and flexibility. On the other hand, middle to large scales of agent systems are not easily built due to the essence of distribution, interaction, human and domain involvement, and openness. In fact, many challenging factors in agent and mining systems are similar or complementary.

(b) Constraint and environment: Both agent and mining systems need to interact with the environment, and tackle the constraints surrounding a system. In agent communities, environment could present characters such as openness, accessibility, uncertainty, diversity, temporality, spatiality, and/or evolutionary and dynamic processes. These factors form varying constraints on agents and agent systems. Similar issues can also be found from real-world data mining, for instance, temporal and spatial data mining. The dynamic business process and logics surrounding data mining make the mining very domain-specific and sensitive to its environment.*² Domain intelligence* Domain intelligence widely surrounds agent and mining systems. Both areas need to understand, define, represent, and involve the roles and components of domain intelligence. In particular, it is essential in agent mining interaction to model domain and prior knowledge, and to involve it to enhance agent-mining intelligence and actionable capability.

(c) Human intelligence: Both agent and mining need to consider the roles and components of human intelligence. Many roles may be better played by humans in agent-mining interaction. To this end, it is necessary to study the definition and major components of human intelligence, and how to involve them in agent mining systems. For instance, mechanisms should be researched on user modeling, user and business friendly interaction interfaces, and communication languages for agent-mining system dialogue.

(d) Knowledge engineering and management: To support the involvement of domain and human intelligence, proper mechanisms of knowledge engineering and management are substantially important. Tasks such as the management, representation, semantic relationships, transformation and mapping between multiple domains, and meta-data and meta-knowledge are essential for involving roles and data/knowledge intelligence in building up agent-mining simians.

(e) Nonfunctional requirements Nonfunctional requests are essential in real-world mining and agent systems. The agent-mining simians may more or less address nonfunctional requirements such as efficiency, effectiveness, action ability, and user and business friendliness.

8. Disciplinary Framework of Agent and Mining Interaction and Integration

This section aims to draw a concept map of agent mining as a scientific field. We observe this from the following perspectives: evolution process and characteristics, agent-mining interaction framework.

8.1 Evolution process and characteristics

As an emerging research area, agent mining experiences the following evolution process, and presents the following unprecedented characteristics.

(a) From one-way interaction to wo-way interaction: The area was originally initiated by incorporating data mining into agent to enhance agent learning [20, 40].Recently, issues in two-way interaction and integration have been broadly studied in different groups.

(b) From single need-driven to mutual needs-driven: Original research work started on the single need to integrate one into the other, whereas it it is now driven by both needs from both parties. As discussed in [12, 8], people have found many issues in each of the related communities. These issues cannot be tackled by simply developing internal techniques. Rather, techniques from other disciplines can greatly complement the problem-solving when they are combined with existing techniques and approaches. This greatly drives the development of agent-driven data mining and data mining-driven agents.

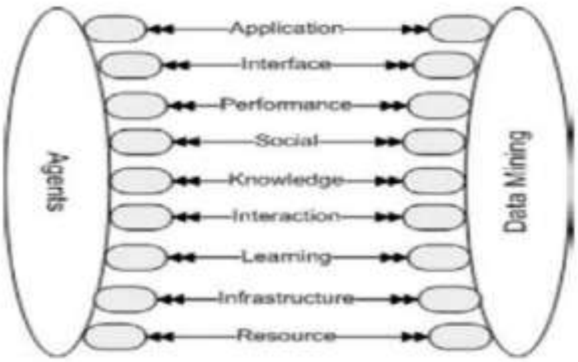
(c) Intrinsic associations and utilities: The interaction and integration between agents and data mining is also driven and connected by intrinsic overlap, associations, complementation and utilities of both parties, as discussed in [5, 6]. This drives the research on mutual issues, and the synergetic research and systems coupling both technologies, into a more advanced form.

(d) Application drives: Application request is one of the key driving forces of this new trend. We present some major application domains and problems that may be better handled by both agent and mining techniques.

(e) Major research groups and researchers [6] :in respective communities tend to undertake both sides of research. Some of them are trying to link them together to solve problems that cannot be tackled by one of them alone, for instance, agent-based distributed learning [30, 31, 32, 25, 26], agent-based data mining infrastructure [4, 5, 26], or data mining driven agent intelligence enhancement [4, 35].

8.2 Agent-mining interaction framework

The interaction and integration between agents and data mining are comprehensive, multiple dimensional, and inter-disciplinary. As an emerging scientific field, *agent mining* studies the methodologies, principles, techniques and applications of the integration and interaction between agents and data mining, as well as the community that focuses on the study of agent mining. On the basis of complementation between agents and data mining, agent mining fosters a synergy between them from different dimensions, for instance, *resource*, *infrastructure*, *learning*, *knowledge*, *interaction*, *interface*, *social*, *application* and *performance*. As shown in Figure 3, we briefly discuss these dimensions.



(a) Resource layer: Interaction and integration may happen on data and information levels;
Infrastructure layer: Interaction and integration may be on infrastructure, architecture and process sides;
(b) Knowledge layer: Interaction and integration may be based on knowledge, including domain knowledge, human expert knowledge, meta-knowledge, and knowledge retrieved, extracted or discovered in resources;

(c) Learning layer: Interaction and integration may be on learning methods, learning capabilities and performance perspectives;

(d) Interaction layer: Interaction and integration may be on coordination, cooperation, negotiation, communication perspectives;

(e) Interface layer: Interaction and integration may be on human-system interface, user modeling and interface design;
(f) Social layer: Interaction and integration may be on social and organizational factors, for instance, human roles;
(g) Application layer: Interaction and integration may be on applications and domain problems;

(h) Performance layer: Interaction and integration may be on the performance enhancement of one side of the technologies or the coupling system.

From these dimensions, many fundamental research issues/problems in agent mining emerge. Correspondingly, we can generate a high-level research map of agent mining as a disciplinary area. Figure .4 shows such a framework, which consists of the following research components: *agent mining foundations*, *agent-driven data processing*, *agent-driven knowledge discovery*, *mining-driven multi-agent systems*, *agent-driven information processing*, *mutual issues in agent mining*, *agent mining systems*, *agent mining applications*, *agent mining knowledge management*, and *agent mining performance evaluation*. We briefly discuss them below.

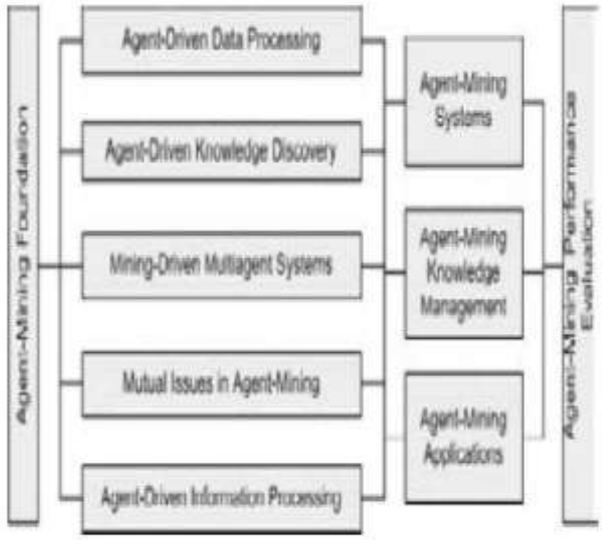


Figure 4: Agent-Mining Disciplinary Framework.

(1) Agent mining foundations: Studies issues such as the challenges and prospects, research map and theoretical underpinnings, theoretical foundations, formal methods, and frameworks, approaches and tools.

(2) Agent-driven data processing: Studies issues including multi-agent data coordination ,multi-agent data extraction, multi-agent data integration, multi-agent data management, multi-agent data monitoring, multi-agent data processing and preparation, multi-agent data query and multi-agent data warehousing;

(3) Agent-driven knowledge discovery: Studies problems like multi-agent data mining infrastructure and architecture, multi-agent data mining process modeling and management, multi-agent data mining project management, multi-agent interactive data mining infrastructure, multi-agent automated data learning, multi-agent cloud computing, multi-agent distributed data mining, multi-agent dynamic mining, multi-agent grid computing, multi-agent interactive data mining, multi-agent online mining, multi-agent mobility mining, multi-agent multiple data source mining, multi-agent

ontology mining, multi-agent parallel data mining, multi agent peer-to-peer mining, multi-agent self-organizing mining, multi-agent text mining, multi-agent visual data mining, and multi-agent web mining; *Mining-driven multi-agent systems (MAS)* studies issues such as data mining driven MAS adaptation, data mining-driven MAS behavior analysis, data mining driven MAS communication, data mining-driven MAS coordination, data mining driven MAS dispatching, data mining-driven MAS distributed learning, data mining-driven MAS evolution, data mining-driven MAS learning, data mining driven MAS negotiation, data mining-driven MAS optimization, data mining driven MAS planning, data mining-driven MAS reasoning, data mining-driven MAS recommendation, data mining-driven MAS reputation/risk/trust analysis, data mining-driven self-organized and self-learning MAS, data mining-driven user modeling and servicing, and semi-supervised MAS learning;

(4) Agent-driven information processing: Multi-agent domain intelligence involvement, multi-agent human-mining cooperation, multi-agent enterprise application integration, multi-agent information gathering/retrieval, multi-agent message passing and sharing, multi-agent pattern analysis, and multi-agent service oriented computing.

(5) Mutual issues in agent mining: Including issues such as actionable capability, constraints, domain knowledge and intelligence, dynamic, online and ad-hoc issues, human role and intelligence, human-system interaction, infrastructure and architecture problems, intelligence meta synthesis, knowledge management, lifecycle and process management, networking and connection, nonfunctional issues, ontology and semantic issues, organizational factors, reliability, reputation, risk, privacy, security and trust, services, social factors, and ubiquitous intelligence; *Agent mining knowledge management:* knowledge management is essential for both agents and data mining, as well as for agent mining. This involves the representation, management and use of ontologies, domain knowledge, human empirical knowledge, meta-data and meta-knowledge, organizational and social factors, and resources in the agent-mining symbionts. In this, formal methods and tools are necessary for modeling, representing and managing knowledge. Such techniques also need to cater for identifying and distributing knowledge, knowledge evolution in agents, and enabling knowledge use.

(6) Agent mining performance evaluation: *R*esearches on methodologies, frameworks, tools and test beds for evaluating the performance of agent mining, and performance benchmarking and metrics. Besides technical performance such as accuracy and statistical significance, business-oriented performance such as cost, benefit and risk are also important in evaluating agent mining. Other aspects such as mobility, reliability, dependability, trust, privacy and reputation, etc., are also important in agent mining.

(7) Agent mining systems: this research component studies the formation of systems, including techniques for the frameworks, modeling, design and software eng.

9. Applications

As we can see from many references, the proposal of agent mining is actually driven by broad and increasing applications. Many researchers are developing agent mining systems and applications dealing with specific business problems and for intelligent information processing. For instance, we summarize the following application domains.

- Artificial immune systems
- Artificial and electronic markets
- Auction
- Business intelligence
- Customer relationship management
- Distributed data extraction and preparation
- E-commerce
- Finance data mining
- Grid computing
- Healthcare
- Internet and network services, e.g., recommendation, personal assistant, searching retrieval, extraction services
- Knowledge management Marketing
- Network intrusion detection
- Parallel computing, e.g., parallel genetic algorithm
- Peer-to-peer computing and service
- Semantic web
- Text mining
- Web mining.

11. Conclusions

Agent and distributed data mining interaction and integration has emerged as a prominent and promising area in recent years. The dialogue between agent technology and data mining can not only handle issues that are hardly coped with in each of the interacted parties, but can also result in innovative and super-intelligent techniques and symbionts much beyond the individual communities.

This chapter presents a high-level overview of the development and major directions in the area. The investigation highlights the following findings: (1) agent mining interaction is emerging as a new area in the scientific family, (2) the interaction is increasingly promoting the progress of agent and mining communities, (3) it results in ever-increasing development of innovative and significant techniques and systems towards super-intelligent symbionts. As a new and emerging area, it has many ope issues waiting for the significant involvement of research resources, in particular practical and research projects from both communities. We believe the research and development on agent mining is very promising and worthy of substantial efforts by both established and new researchers.

12. References

[1]. Aciar, S., Zhang, D., Simoff, S., and Debenham, J.: Informed Recommender Agent: Utilizing Consumer Product Reviews through Text Mining. Proceedings of IADM2006. IEEE Computer Society (2006)

[2]. Batik's., Cho, J., and Bala, J.: Performance Evaluation of an Agent Based Distributed Data Mining System. Advances in Artificial Intelligence, Volume 3501/2005 (2005)

[3]. Cory, J., Butz, Nguyen, N., Takama, Y., Cheung, W., and Cheung, Y.: Proceedings of IADM2006 (Chaired by Longbing Cao, Zili Zhang, Vladimir Samoilov) in WI-IAT2006 Workshop Proceedings. IEEE Computer Society (2006)

[4]. Cao, L., Wang, J., Lin, l., and Zhang, C.: Agent Services-Based Infrastructure for Online Assessment of Trading Strategies. Proceedings of IAT'04, 345-349 (2004).

[5]. Cao, L.: Integration of Agents and Data Mining. Technical report, 25 June 2005. http://wwwstaff.it.uts.edu.au/ lbcao/publication/publications.htm.

[6]. Cao, L., Luo, C. and Zhang, C.: Agent-Mining Interaction: An Emerging Area. AIS-ADM, 60-73 (2007).

[7]. Cao, L., Luo, D., Xiao, Y. and Zheng, Z. Agent Collaboration for Multiple Trading Strategy Integration. KES-AMSTA, 361-370 (2008).

[8]. Cao, L.: Agent-Mining Interaction and Integration – Topics of Research and Development. http://www.agentmining.org/

[9]. Cao, L.: Data Mining and Multiagent Integration. Springer (2009).

[10]. Cao, L. and Zhang, C. F-trade: An Agent-Mining Symbiont for Financial Services. AAMAS 262 (2007).

[11]. Cao, L., Yu, P., Zhang, C. and Zhao, Y. Domain Driven Data Mining. Springer (2009).

[12]. Cao, L., Gorodetsky, V. and Mitkas, P. Agent Mining: The Synergy of Agents and Data Mining. IEEE Intelligent Systems (2009).

[13]. Cao, L. Integrating Agent, Service and Organizational Computing. International Journal of Software Engineering and Knowledge Engineering, 18(5): 573-596 (2008)

[14]. Cao, L. and He, T. Developing Actionable Trading Agents. Knowledge and Information Systems: An International Journal, 18(2): 183-198 (2009).

[15]. Cao, L. Developing Actionable Trading Strategies, Knowledge Processing and Decision Making in Agent-Based Systems, 193-215, Springer (2008).

[16]. Cao, L., Zhang, Z., Gorodetsky, V. and Zhang, C.. Editor's Introduction: Interaction between Agents and Data Mining, International Journal of Intelligent Information and Database Systems, Inderscience, 2(1): 1-5 (2008).

[17]. Cao, L., Gorodetsky, V. and Mitkas, P. Editorial: Agents and Data Mining. IEEE Intelligent Systems (2009).

[18]. Cao, L. Agent & Data Mining Interaction, Tutorial for 2007 IEEE/WIC/ACM Joint Conferences on Web Intelligence and Intelligent Agent Technology (2007).

[19]. Cao, L., Zhang, C. and Zhang, Z. Agents and Data Mining: Interaction and Integration, Taylor & Francis (2010).

[20]. Brazdil, P., and Muggleton, S.: Learning to Relate Terms in a Multiple Agent Environment.EWSL91 (1991)

[21]. Davies, W.: ANIMALS: A Distributed, Heterogeneous Multi-Agent Learning System. MSc Thesis, University of Aberdeen (1993)

[22]. Davies, W.: Agent-Based Data-Mining (1994)

[23]. Edwards, P., and Davies, W.: A Heterogeneous Multi-Agent Learning System. In Deen, S.M. (ed) Proceedings of the Special Interest Group on Cooperating Knowledge Based Systems. University of Keele (1993) 163-184.

[24]. Gorodetsky, V., Liu, J., Skormin, V. A.: Autonomous Intelligent Systems: Agents and Data Mining book. Lecture Notes in Computer Science Volume 3505 (2005)

[25]. Gorodetsky, V.; Karsaev, O.and Samoilov, V.: Multi-Agent Technology for Distributed Data Mining and Classification. IAT 2003. (2003) 438 - 441

[26]. Gorodetsky, V., Karsaev, O. and Samoilov, V.: Infrastructural Issues for Agent-Based Distributed Learning. Proceedings of IADM2006, IEEE Computer Society Press

[27]. Han, J., and Kamber, M.: Data Mining: Concepts and Techniques (2nd version). Morgan Kaufmann (2006)

[28]. Kaya, M. and Alhajj, R.: A Novel Approach to Multi-Agent Reinforcement Learning: Utilizing OLAP Mining in the Learning Process. IEEE Transactions on Systems, Man and Cybernetics, Part C, Volume 35, Issue 4 (2005) 582 - 590

[29]. Kaya, M. and Alhajj, R.: Fuzzy OLAP Association Rules Mining-Based Modular Reinforcement Learning Approach for Multi-Agent Systems. IEEE Transactions on Systems, Man and Cybernetics, Part B, Volume 35, Issue 2, (2005) 326 - 338

[30]. Klusch, M., Lodi, S. and Gianluca, M.: The Role of Agents in Distributed Data Mining: Issues and Benefits. Intelligent Agent Technology (2003): 211 - 217

[31]. Klusch, M., Lodi, S. and Moro,G.: Agent-Based Distributed Data Mining: The KDEC Scheme. Intelligent Information Agents: The AgentLink Perspective Volume 2586 (2003) Lecture Notes in Computer Science

[32]. Klusch, M., Lodi, S. and Moro, G.: Issues of Agent-Based Distributed Data Mining. Proceedings of AAMAS, ACM Press (2003)