Chapter 1
Introduction to Agent Mining Interaction and Integration

Longbing Cao

Abstract In recent years, more and more researchers have been involved in research on both agent technology and 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 data mining. We refer this to *agent mining* as a new area. The marriage of agents and data mining is driven by challenges faced by both communities, and the need of developing more advanced intelligence, information processing and systems. This chapter presents an overall picture of agent mining from the perspective of positioning it as an emerging area. We summarize the main driving forces, complementary essence, disciplinary framework, applications, case studies, and trends and directions, as well as brief observation on agent-driven data mining, 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 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.

1.1 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.

Longbing Cao
Faculty of Engineering and Information Technology, University of Technology Sydney, Australia, e-mail: lbcao@it.uts.edu.au

L. Cao (ed.), Data Mining and Multi-agent Integration, DOI: 10.1007/978-1-4419-0522-2_1,

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, e-commerce, 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]. This edited book, as the first one on this exciting topic, once again evidences the strong need and potential of agent-mining interaction and integration (*agent mining* for short).

This chapter presents an overall picture of this emerging field, data mining and multi-agent integration. We first analyze the respective and common challenges in agents and 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, data mining-driven agents, and mutual issues in agent mining. Applications and open issues are then 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.

1.2 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. Fig. 1.1 illustrates these challenges. 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.

1.2.1 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. In the following analysis, we concern ourselves with the challenges that may benefit from the involvement of data mining. We explain this from the follow-

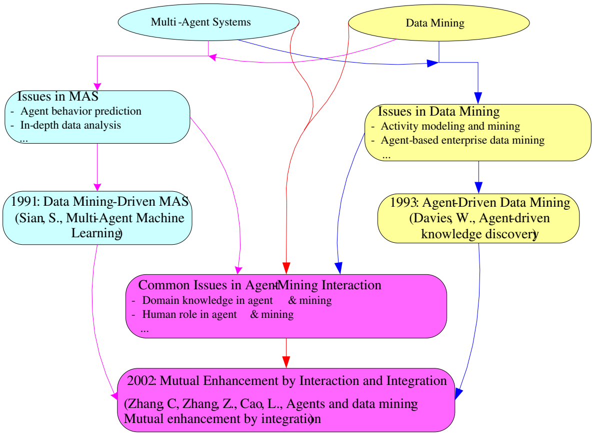

Fig. 1.1: Challenges in agents and data mining

ing aspects: agent awareness, agent learning, agent actionability, agent distributed processing, agent in-depth services, and agent constraint processing.

- *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.
- *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 learn-

ing presents either a passive or active manner. In a broad sense, learning can be in a supervised, unsupervised or hybrid manner.

- *Agent actionability* Agent actionability 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.
- *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.
- *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.
- *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, actionability, 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.

1.2.2 Challenges in 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.

- *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.
- *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.
- *Supporting parallel and distributed mining* One of the major efforts of data mining research is to enhance the performance of data mining algorithms. This is usually conducted through designing efficient data structures and computational methods to reduce computational complexities. In many cases, computational performance can be greatly improved through developing parallel algorithms. In other cases, distributed computing is necessary such as dealing with distributed data sources or applications, or peer-to-peer computing is required. However, how to design effective and efficient parallel and distributed algorithms is an issue.
- *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.
- *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 unadaptable. 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 automated and adaptive intelligent techniques.

- *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.

1.2.3 Mutual challenges in agent and 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. Fig. 1.1 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.

- *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.
- *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 surround-

ing data mining make the mining very domain-specific and sensitive to its environment.

- *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.
- *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.
- *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 symbionts.
- *Nonfunctional requirements* Nonfunctional requests are essential in real-world mining and agent systems. The agent-mining symbionts may more or less address nonfunctional requirements such as efficiency, effectiveness, actionability, user and business friendliness.

1.3 Complementary essence and interaction potential of agents and data mining

Why do the above challenges matter to both sides of agents and data mining? Why is the interaction and integration between agents and data mining important? There are both explicit and implicit reasons. Explicit reasons may include the following system complexities.

- Explicit limitations and challenges in pure agent systems, as addressed in [5, 6, 7], can be complemented by data mining, for instance, data mining driving agent learning, user modeling and information analysis.
- Explicit limitations and challenges in pure data mining systems, as discussed in [5, 6, 7], can be better serviced by agent technology, for instance, agent-based data mining infrastructure, agents for data management and preparation, agent-based service provision.
- The integration of agents and data mining has the potential to result in new strengths and advantages that cannot be delivered by any single side, for instance,

leading to more intelligent agent-mining symbiont fusing capabilities of in-depth perception, learning, adaptation, discovery, reasoning, and decision.

Implicit driving forces for including the above mutual issues are equally significant.

- Agent-mining symbionts are substantially essential for dealing with complicated intelligence phenomena and system complexities in complex intelligent systems. Simple intelligent systems and other issues that can be tackled using one side of these technologies, for instance, an agent-based data integration system, may not necessarily involve both sides.
- The emergence of intelligence in agent-mining interaction may massively strengthen the problem-solving capability of an intelligent system, which cannot be carried out by either part.
- Implicit roles need to be discovered through interdisciplinary studies, which may extensively promote either one side or the whole of an agent-mining integrative system, once the roles are disclosed and properly developed.
- New research issues, opportunities, techniques and systems may be triggered in the agent mining community.

It is arguable that agents and mining are complementary. The agent-mining interaction can enhance both sides considerably through introducing new approaches and techniques to solve those domain-specific challenges that cannot be tackled well by either methods. Some typical benefits and roles in agent and mining areas that can be achieved through agent-mining interaction.

- Enhancing agents through data mining. Agent-mining interaction was originally initiated by data mining driven agent learning in 1991 [20, 40]. Data mining has the potential to enhance agent technology through introducing and improving the learning and reasoning capability of agents. Agents can be enhanced through involving data mining in broad aspects, in particular, agent learning, agent coordination and planning, user modeling and servicing, and network servicing.
- Promoting data mining through agent. Sometime around 1993, another effort was started on agent-based data mining [21, 22, 23], namely to utilize agent technology to enhance data mining. The enhancement may be embodied in terms of varying aspects, for instance, agent-based KDD infrastructure, agent-based distributed processing, agent-based interactive data mining, and agent-based data warehouse.
- Building super intelligent symbionts. As evidenced by the agent service-based trading support system F-Trade [10], the use of agent mining can lead to more intelligent systems that can best fuse the strengths of agents in building intelligent systems as well as the beauty of data mining in processing deep knowledge.

1.4 A 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, and theoretical underpinnings for agent mining.

1.4.1 Evolution process and characteristics

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

- *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.
- *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.
- *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.
- *Application drives*: Application request is one of the key driving forces of this new trend. In Section 1.8.1, we present some major application domains and problems that may be better handled by both agent and mining techniques.
- *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].
- Broad research covering *theoretical, technological and practical perspectives*: Publications and projects have involved not only technological issues, but also theoretical and practical problems. A cross-disciplinary and multi-dimensional study roadmap is clear.

We also draw an evolutionary tree of this area by combining the emergence of significant landmarks and events in the life of agent mining (see Fig. 1.4.1).

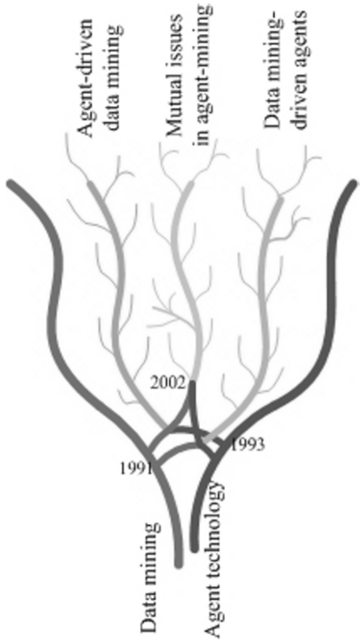

Fig. 1.2: Evolution of agent mining as a scientific area

As identified in a recent position meeting and related activities[5], there are many research topics and open issues from both sides of agent and mining interaction. In particular, issues for agent-driven data mining, and issues for mining-driven agents are attracting research interest. However, there are some mutually fundamental issues that are not receiving attention in the emerging research. These issues are significant because of their fundamental and necessary roles in establishing a symbiotic relation between agents and mining.

Through reviewing the related work in the above areas, there is a clear indication that agent-mining interaction and integration has emerged as a prominent, challenging, dynamic and exciting area. It evidences that

1. Agent-mining interaction is attracting ever-increasing attention from both agents and data mining communities,
2. The interaction and integration between agent and mining can greatly complement and strengthen each side of both communities. Some complicated challenges in either community may be effectively and efficiently tackled through agent-mining interaction,

3. Furthermore, as a newly emergent area, agent and mining interaction and integration has the potential to create new interesting symbiosis opportunities in both academic and business worlds,
4. As a new open area, however, there are many issues awaiting research and development from theoretical, technological and practical perspectives.

1.4.2 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 Fig. 1.3, we briefly discuss these dimensions.

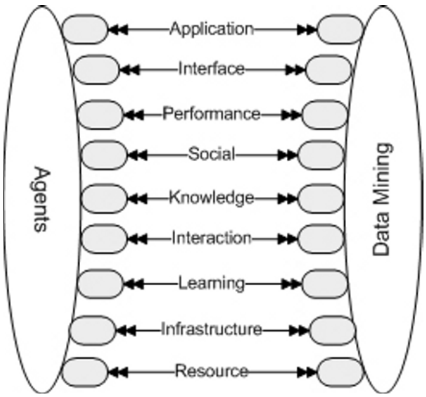

Fig. 1.3: Multi-Dimensional Agent-Mining Synergy.

- 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;
- 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;

- Learning layer – interaction and integration may be on learning methods, learning capabilities and performance perspectives;
- Interaction layer – interaction and integration may be on coordination, cooperation, negotiation, communication perspectives;
- Interface layer – interaction and integration may be on human-system interface, user modeling and interface design;
- Social layer – interaction and integration may be on social and organizational factors, for instance, human roles;
- Application layer – interaction and integration may be on applications and domain problems;
- 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 1.4 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.

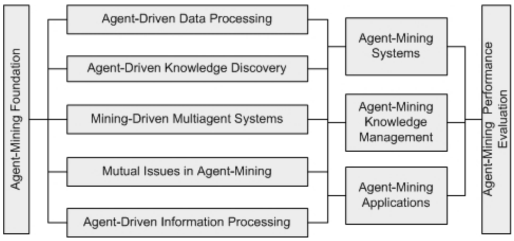

Fig. 1.4: Agent-Mining Disciplinary Framework.

- *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;
- *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;
- *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;
- *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;
- *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 metasynthesis, 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.
- *Agent mining performance evaluation* researches on methodologies, frameworks, tools and testbeds 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.
- *Agent mining systems:* this research component studies the formation of systems, including techniques for the frameworks, modeling, design and software engi-

neering of agent-mining systems. It provides agents and data mining technologies as basic resources, thus they can perform as parts of the system. Techniques and tools for engineering and constructing agent-mining systems are important for handling various kinds of applications. A specific agent-mining system may be extended to fit into some applications based on the provided features of the system. With regards to integrated systems, applications can be constructed from the pre-defined framework of a particular problem domain or fully tailored to serve the purpose of the applications. Either way, system implementation and deployment are to be further investigated since development of the two technologies has been done in parallel. It is hard to generate dedicated platforms for agent mining systems.
- *Agent mining applications* This refers to any real-world applications and domain problems that can be better handled by agent mining technologies. Based on the need from particular applications, any issues discussed in the above topics may be engaged here. For instance, in some cases, an agent mining simulation system needs to be built for us to understand the working mechanism and potential optimization of a complex social network. In other cases, the enhancement of learning capability is the main task, and appropriate learning tools need to be used on demand.

1.4.3 Theoretical underpinnings for agent mining

Current research on agent mining mainly focuses on the application of agents in data mining or data mining in agents. No systematic work has been conducted on developing foundations for agent mining. In fact, we believe it is very important to foster research on fundamental issues such as:

- how to synergize agents and data mining.
- what methodologies are needed to synergize agents and data mining?
- what are the lifecycle, process and outcome aspects to synergizing agents and data mining?

This needs to be investigated from methodological, technical and tool perspectives. For this, we need to investigate the theoretical underpinnings for agent mining.

Obviously, to support agent mining, many disciplines related to either agents or data mining need have been involved, for instance, artificial intelligence, machine learning, and logic. We form the following multiple layers of framework for developing an agent mining discipline: (1) Theoretical foundation, (2) Fundamental technologies, and (3) Supporting techniques and tools.

From the *theoretical foundation* perspective, agent mining draws theoretical support from multiple disciplines, including mathematics, logics, information sciences, intelligence sciences, system sciences, cognitive sciences, many particular disciplines in social sciences such as business, and behavioral sciences. Mathematics and logics provide formal methods for learning, modeling and knowledge representation, reasoning, discovery, transformation and presentation. Information and intelligence sciences provide support for intelligent information processing and systems in agent mining. System sciences furnish methodologies and techniques for the understanding, modeling, simulation, deployment and system integration. Cognitive sciences incorporate principles and methods for understanding human behavior belief, and the intention and goal of human behavior, in order to involve human intelligence in the agent mining process, supporting user modeling and services, and human-centered computing. The social sciences supply the foundations for conceiving organizational and social factors and business processes surrounding domain problems using agent mining. Many areas may be involved, for instance, economics, finance, business, behavior science, choice, and social network analysis, which are important for understanding environmental impact, interaction, and requirements on deliverables.

Fundamental technologies may involve many aspects in information and technology fields, for instance, user modeling, formal methods, logics, knowledge engineering, data engineering, ontological engineering, semantic web, machine learning, artificial intelligence, software engineering, information systems, and interaction design. Specific techniques in risk management and analysis, organizational theory, sociology, psychology, economics and finance are important for incorporating organizational, social and domain factors into agent mining. Emerging areas such as organizational computing, social computing, ubiquitous computing and collective intelligence can contribute to the agent mining family as well. In fact, agent mining will involve a majority of disciplines and a body of knowledge in the modern science and technology family.

To make agent mining work, many *supporting techniques and tools* are essential. This may involve effective techniques and tools for representing, modeling, analyzing, presenting, integrating and employing agents and data mining. In particular, techniques and tools are required to support information access, fusion, processing, discovery, user interface design, visualization, dynamic and adaptive interaction, information and knowledge sharing, integration and management, and mechanisms for supporting the performance evaluation and improvement of agent mining bodies. For instance, how can we evaluate the performance of agent mining symbionts? In data mining-driven agent planning, how can we involve data mining findings to make agent planning smarter, more stable and predictable?

1.4.4 Agent mining lifecycle and process

In general, it is very hard to define or extract a generic lifecycle and process for agent mining interaction and integration, as has been done for agent system development and data mining. This is partially due to the diversity of methodologies, techniques and community interests in both agents and data mining fields. In fact, the lifecyle and process for any concrete case study or system has to be specific. However, this does not mean this has nothing to do with a general understanding of lifecyle and process for integrating agents and data mining.

If agents are mainly used to support data mining, the general lifecycle and process would fit into that of data mining and knowledge discovery, from data extraction, transformation, loading, to modeling, initial evaluation, refinement, final improvement and delivery of findings. On the other hand, if data mining is mainly used for agent system construction, a likely process is to follow the software engineering lifecycle, such as requirement analysis, system analysis, database and semantic design, architectural and detailed system design, user interface design, to system implementation, performance evaluation and refinement. In either case, the lifecycle and process need to be customized and enhanced toward catering for advantages and requests from the other end of the cycle; for instance, the data mining process needs to be supported in constructing a data mining-driven agent system.

In other cases, mutual issues are focused by involving both agents and data mining, for instance, research on the involvement of domain knowledge and intelligence into agent mining symbionts. The processes for such a case would more likely follow that usually used for handling the issue by considering the special needs from agents and data mining aspects.

However, due to the extra contributions and needs from the involvement of the other technique, it is interesting to develop a customizable lifecycle and process for agent-driven data mining and data mining driven agent systems. For instance, if data mining is used to strengthen agent intelligence, how should data mining agents be trained and tested? Extra procedures may be required for such data mining agents.

1.5 Agent-Driven Distributed Data Mining

This section particularly discusses the state-of-the-art of agent-driven knowledge discovery (otherwise known as multi-agent-driven data mining, multi-agent data mining) [12]. As discussed in the above, agent-driven knowledge discovery forms a big area for agent mining. It is actually the mostly addressed area since the proposal of the integration between agents and data mining.

1.5.1 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.

1.5.2 The needs of agent-driven distributed data mining

The practical implementation of distributed and P2P data mining and machine learning creates many new challenges. While analyzing these challenges, [30] argues why agent technology is best able to cope with them in terms of autonomy, interaction, dynamic selection and gathering, scalability, multi-strategy and collaboration. Other reasons include privacy, mobility, time constraints (stream data which is too late to extract and then mine), and computational costs and performance requests.

- Isolation of data sources. Distributed and multiple data sources are often isolated from each other. For in-depth understanding of a business problem, it is essential to bring relevant data together through centralized integration or localized communication. From this, agent planning and collaboration, mobile agents, agent communication and negotiation can benefit.
- Mobility of source data and computational devices. Data and device mobility requires the perception and action of data mining algorithms on a mobile basis. Mobile agents can adapt to mobility very well.
- Interactive DDM. Pro-actively assisting agent is necessary to drastically limit how much the user has to supervise and interfere with running the data mining process.

- Dynamic selection of sources and data gathering. One challenge for an intelligent data mining agent acting in an open distributed environment, in which to pursue the DM tasks, for example, where the availability of data sites and their content may change at any time, is to discover and select relevant sources. In these settings, DM agents may be applied to adaptively select data sources according to given criteria such as the expected amount, type and quality at the considered source, actual network and DM server load. Agents may be used, for example, to dynamically control and manage the process of data gathering.
- Time constraints on distributed data sources. Some data distributed in different storages is dependent on time, e.g., time differences.
- Multi-strategy DDM. For some complex application settings, an appropriate combination of multiple data mining techniques may be more beneficial than the application of a particular one. DM agents may learn in, due course which of their deliberative actions to choose, depending on the type of data retrieved from different sites and the mining tasks to be pursued.
- Collaborative DDM. DM agents may operate independently on data they have gathered at local sites and then combine their respective models. Alternatively, they may agree to share potential knowledge as it is discovered, in order to benefit from the additional options of other DM agents.
- Privacy of source data. Distributed local data is not allowed to be extracted and integrated with other sources directly, due to privacy issues. A DM agent with authority to access and process the data locally can dispatch identified local patterns for further engagement with findings from other sources.
- Organizational constraint on distributed data sources. In some organizations, business logic, process and work-flow determine the order of data storage and access. This, therefore, augments the complexity of DDM. Agents located in each storage area can communicate with each other and dispatch the DDM algorithm agents instantly, once the response is over.

1.5.3 Research issues in agent driven data mining

There are many open issues in the research direction of agent driven data mining. In establishing an agent-based enterprise data mining infrastructure, one may study organization and society-oriented study system analysis and design techniques for large-scale agent systems. Correspondingly, solutions for agent service based application integration, distributed data preparation, distributed agent coordination and parallel agent computing should be considered. In many cases of data mining, people should study algorithms that can adapt to dynamic data changes, dynamic user requests. To this end, it has the potential for agents to detect and reason such changes. Automated and adaptive data mining algorithms should be studied. The following is a list of some research open issues and promising areas.

- Activity modeling and mining
- Agent-based enterprise data mining

- Agent-based data mining infrastructure
- Agent-based data warehouse
- Agent-based mining process and project management
- Agent-based distributed data mining
- Agent-based distributed learning
- Agent-based grid computing
- Agent-based human mining cooperation
- Agent-based link mining
- Agent-based multi-data source mining
- Agent-based interactive data mining
- Agent-enriched ontology mining
- Agent-based parallel data mining
- Agent-based web mining
- Agent-based text mining
- Agent-based ubiquitous data mining
- Agent knowledge management in distributed data mining
- Agent for data mining data preparation
- Agent-human-cooperated data mining
- Agent networks in distributed knowledge discovery and servicing
- Agent service-based KDD infrastructure
- Agent-supported domain knowledge involvement in KDD
- Agent system providing data mining services
- Automated data mining learning
- Autonomous learning
- Distributed agent-based data preprocessing
- Distributed learning
- Domain intelligence in agent-based data mining
- Mobile agent-based knowledge discovery
- Protocols for agent-based data mining
- Self-organizing data mining learning.

1.6 Data Mining-Driven Agents

This section discusses data mining-driven multi-agent systems [12]. In contrast to the previous section, this section emphasizes agents empowered by more informative knowledge provided by data mining.

1.6.1 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 efficiency.

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 DM-extracted 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.

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 agent-mediated 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 crawling with clusters of synergetic crawler agents.

1.6.2 The needs of data mining-driven agents

Data mining-driven multi-agent systems present attractive features to create more intelligent systems as discussed below:

- The combination of autonomy (MAS) and knowledge (DM) provides adaptable systems. Knowledge discovered in data and then fed into agents can greatly enhance the self-organization and learning performance of agents.
- DM can greatly enhance the learning and knowledge processing capability of agents, through involving DM algorithms in the building-blocks of agent learning systems. As a result, agents can learn from the data and from the environment before planning and reasoning are made.
- DM can enhance the agent capability of handling uncertainty through historical event analysis and dynamic mining and active learning; through mining agent behavioral data, it is possible to reach a balance and trade-off between agent autonomy and supervised evolution; the outcomes of self-organization and emergence become much more certain, controllable and predictable.
- The rigidity and lack of exploration of deductive reasoning systems is overcome. Rules are no longer hard-coded into systems and their modification is only a matter of retraining.
- DM techniques such as association rule extraction have no equivalent in agent systems. These techniques now provide agents with the ability of learning, discovery, probing and searching.

- Real-world databases often contain missing, erroneous data and/or outliers. Through clustering, noisy logs are assimilated and become a part of a greater group, smoothing down differences, while outliers are detected and rejected. Through classification, ambiguous data records can be validated and missing data records can be estimated. Rule-based systems cannot handle such data efficiently without increasing their knowledge-base and therefore their maintenance cost.
- The presented approach favors the combination of inductive and deductive reasoning models. In some of the demonstrators presented, there were agents deploying deductive reasoning models ensuring, thereby ensuring system soundness. Nevertheless, these agents decide on data already preprocessed by inductive agents. In this way, the dynamic nature of the application domains is satisfied, while the set of deductive results (knowledge-bases of deductive agents) becomes more compressed and robust.
- Even though the patterns and rules generated through data mining cannot be defined as *sound*, there are metrics deployed to evaluate the performance of the algorithms. Total mean square error (clustering), support-confidence (association rules), classifier accuracy (classification), and classifier evaluation (genetic algorithms), among others, are employed for evaluating the knowledge models extracted. The engagement of data mining performance evaluation into agents satisfies the need of agents for knowledge quality and model evaluation, model testing and model comprehension.
- Usually DM tools are introduced to enterprises as components-off-the-shelf. These tools are used by human experts to examine their corporate or environmental databases to develop strategies and take decisions. This procedure often proves time-consuming and inefficient. By exploiting concurrency and multiple instantiation of agent types (cloning capabilities) of MAS systems, and by applying data mining techniques for embedding intelligent reasoning into them, useful recommendations can be much more quickly diffused while parallelism can be applied to non-related tasks, pushing system performance even higher.

1.6.3 Research issues in data mining driven agents

Research issues on the involvement of data mining and knowledge discovery in agents can be in varying aspects. Appropriate learning models need to be studied and involved to support reasoning, adaptation and evolution in multi-agents. New communication, planning, coordination and dispatching mechanisms may be developed by discovering the interaction of proper patterns and relationships between agents. In human-agent interaction, we need to develop proper algorithms to discover user behavior patterns, so that agent-user interaction and servicing can be more effective. In distributed conditions, we need to develop distributed learning algorithms to manage the coordination of agent crossing multiple applications. In the following, we list some topics that are promising areas in data mining driven agent research.

- Collaborative learning in multi-agents
- Data mining-driven agent learning, reasoning, adaptation and evolution
- Data mining-driven multi-agent communication, planning and dispatching
- Data mining-driven user modeling
- Data mining-driven user servicing
- Data mining-driven network servicing
- Data mining-driven agent recommender
- Data mining-driven trading agents
- Data mining agent assistant
- Multi-agent reinforcement learning
- Multi-agent knowledge discovery
- Data mining enhancing agent intelligence enhancement
- Decentralized clustering in large multi-agent systems
- Distributed learning in agent coordination
- Distributed learning in multi-agent systems
- Emergent agent organization and behavior
- Information gathering agents
- Learning agents
- Web mining agents
- Self-learning agents.

1.7 Mutual Issues in Agent Mining

There are many mutual enhancement issues in both agents and data mining respectively, and during the integration. Typical issues consist of architecture and infrastructure problems, actionable capability of agent-mining symbionts, constraints in agents and mining, data intelligence in agent and mining, domain knowledge in agents and mining, evaluation issues such as technical significance and business expectation, gaps filling between technical and business expectations, human intelligence and roles in agents and mining, knowledge management in agents and mining, meta-data and meta-knowledge in agents and mining, usability, expandability, openness, organizational and social factors and issues such as business factors, processes, security, privacy, trust; services request, response, recommendation and management, and finally the meta-synthesis of relevant ubiquitous factors into an effective and integrative system for agent and data mining problem-solving. We now briefly discuss some of these issues.

1.7.1 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, workflows, 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 problem-solving. 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. Bodies of knowledge that may be useful consist of subjects such as cognitive science, human-machine interaction, interaction and interface design, knowledge engineering and management, ontological engineering, evaluation systems, organizational computing, social computing, artificial intelligence, and machine learning.

1.7.2 Mutual research issues in agent mining

After years of unorganized development of the one-way effect discussed above, there is further recolonization of fundamental mutual issues in agent-mining interaction [4, 8, 45, 2], which involve common issues of both parties. The studies on these mutual issues should not only tackle problems of one-way enhancement as discussed in Sections 1.5 and 1.6, but also two-way strengthening in building a super-intelligent agent-mining symbiont. However, these issues have not attracted sufficient attention in the community.

- Architecture and infrastructure problems
- Actionable capability of agent-mining symbionts
- Constraints in agent and mining
- Data intelligence in agent and mining
- Domain knowledge in agent and mining
- Domain intelligence in agent and mining
- Evaluation issues such as technical significance and business expectation
- Gap filling between technical and business expectations
- Human intelligence and role in agent and mining
- Human-system interaction
- Intelligence meta-synthesis in agent and mining
- Knowledge management in agent and mining
- Meta-data and meta-knowledge in agent and mining
- Nonfunctional issues such as usability, expandability, openness
- Ontology issues in agent and mining
- Organizational issues such as business factors, process
- Performance issues such as effectiveness, efficiency, scalability
- Social issues such as security, privacy, trust
- Services request and response, service-oriented management
- System management.

1.8 Applications and case studies

1.8.1 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 services
- Semantic web
- Social intelligence & social network analysis
- Supply chain management
- Trading agents
- Trading optimization and support
- Text mining
- Web mining.

1.8.2 Case Studies: Developing Actionable Trading Agents

The task of *designing smart trading agents* is to endow trading agents with the capabilities of searching strategies in a constrained market environment to satisfy trader preference. We introduce two approaches to designing smart trading agents. One is to design *evolutionary trading agents* [14], which are equipped with evolutionary computing capabilities, and can search strategies from a large candidate strategy space targeting higher *benefit-cost ratio*. The other is to integrate optimal instances from multiple classes of trading strategies into one combined powerful strategy through *collaborative trading agents* [15].

Evolutionary Trading Agents for Parameter Optimization

Evolutionary trading agents have capabilities of evolutionary search computing. They can search trading strategies based on given optimization fitness and specified optimization objectives. Their roles consist of optimization requests (including base strategies and arguments), creating strategy candidates (namely chromosomes), evaluating strategy candidates, crossing over candidate strategies, mutating candidate strategies, re-evaluating candidate strategies, filtering optimal strategies, and so on.

The strategy optimization using evolutionary trading agents is as follows. A *User Agent* receives optimization requests from user-agent interaction interfaces. It forwards the request to *Coordinator Agents* and the *Coordinator Agents* check the availability and validity of the optimized *Strategy Agent* class with *strategyClassID*. If a *Strategy Agent* class is available and optimizable, *Coordinator Agents* call the *Evolutionary Agents* to perform corresponding roles, for instance, *createStrate-*

gyCandidates, *evaluateStrategyCandidates*, *crossoverCandidateStrategies*, *mutateCandidateStrategies*, *re-evaluateCandidateStrategies*, or *returnOptimalStrategies* to optimize the strategy. After the optimization process, *Evolutionary Agents* return *Coordinator Agents* the searched optimal *Strategy Agent* with *strategyID* and corresponding parameter values. *Coordinator Agents* further call the *User Agents* to present the results to traders by invoking *Presentation Agents*.

Trading Agent Collaboration for Strategy Integration

In real-life trading, trading strategies can be categorized into many classes. To financial experts, different classes of trading strategies indicate varying principles of market models and mechanisms. A trading agent often takes a series of positions generated by a specific trading strategy, which instantiates a trading strategy class. Our idea is for multiple trading agents to collaborate with each other and take concurrent positions created by multiple trading strategies to take advantage of varying strategies.

The working mechanism of *trading agent collaboration for strategy integration* is as follows. There are a few *Representative Trading Agents* in the market. Each *Representative Agent* invokes an *Evolutionary Agent* to search for an optimal *Strategy Agent* from a strategy class. *Coordinator Agents* then negotiate with these *Representative Agents* and *Evolutionary Agents* to integrate the identified optimal strategies of *Strategy Agents*. An appropriate integration method is negotiated and chosen by *Coordinator Agents*, *Representative Agents* and *Evolutionary Agents* based on a globally optimal objective.

1.8.3 F-Trade: An agent-mining symbiont

In this section, we briefly introduce an agent-mining symbiont – F-Trade[1] to illustrate the development and use of agent-mining interaction technology in tackling both research and business issues.

F-Trade [4] is the acronym of Financial Trading Rules Automated Development and Evaluation, a web-based automated enterprise infrastructure for trading strategies and data mining on stock/capital markets. The system offers data connection, management and processing services. F-Trade supports online automated plug and play, and automatic input/output interface construction for trading signals/rules and data mining algorithms, data sources, and system components. It provides powerful and flexible supports for online backtesting, training/test, optimization and evaluation of trading strategies and data mining algorithms. Users can plugin, subscribe, supervise and optimize trading strategies and data mining algorithms in a human-machine cooperated manner.

F-Trade is built in Java agent services on top of Windows/Linux/Unix. XML is used for system configuration and metadata management. A super-server functions

[1] www.f-trade.info, www.ftrade.info

as the application server, another one acts as the data warehouse. It is constructed with online connectivity to distributed data sources as well as user-specific data sources.

Major roles played by agents in F-Trade consist of agent service-based architecture, agent-driven human interaction, agent for data source management, data collection and dispatch agents roaming to remote data sources, agentized trading strategies and data mining algorithms, agent and service recommender providing optimum algorithms and rules to users, and so on. Data mining assists the system in aspects such as data mining-driven trading rule/algorithm recommender agents, data mining-driven user services, data mining-driven trading agent optimizers, mining actionable trading rules in generic trading pattern set, parameter tuning of algorithm agents through data mining, etc. Mutual issues involve ontology-based domain knowledge representation and transformation to problem-solving terminology, human involvement and agent-based human interaction with algorithms and the system for algorithm supervision, optimization and evaluation, among others.

1.8.4 Agent Academy

Agent Academy [41] is an open-source framework and integrated development environment (IDE) for creating software agents and multi-agent systems, and for augmenting agent intelligence through data mining. The core objectives of Agent Academy are to:

- - Provide an easy-to-use tool for building agents, multi-agent systems and agent communities.
- - Exploit Data Mining techniques for dynamically improving the behavior of agents and the decision-making process in multi-agent systems.
- - Serve as a benchmark for the systematic study of agent intelligence generated by training them on available information and retraining them whenever needed.
- - Empower enterprise agent solutions by improving the quality of provided services.

Agent Academy has been implemented upon the JADE and WEKA APIs (Application Programming Interfaces) and its second version is now available at Sourceforge [2]. The current release contains 237 Java source files and is spun on over 28,000 lines of code.

Agent Academy proposes a new line of actions for creating DM-enhanced agents, through a simple and well-defined workflow. In a typical scenario the developer must follow a specified methodology in order to build a new agent application. First, agent behaviors, as orthogonal as possible, are built while, in a parallel process, data mining is performed in order to generate knowledge models for the application under development. Next, everything is organized/assigned to agents which, eventually, will constitute the multi-agent system.

[2] http://sourceforge.net/projects/agentacademy

1.9 Trends and directions

As a newly emerging area, agent-mining interaction has an expansive future. Unprecedented potential is embodied through unlimited research interest points in areas of (1) data mining driven agent, (2) agent driven data mining, and (3) mutual issues in agent-mining interaction.

Research and exploitation are advancing to develop the better, tighter, and more organized integration for the systems. We raise some issues here that researchers are encouraged to pursue [12].

- Foundations, including methodologies, formal tools and processes for supporting integration of agents and data mining from multiple dimensions as outlined in Fig. 1.3
- Formulation of formal methodologies, languages, notations, for agent mining software engineering
- Methodologies and techniques supporting effective involvement and integration of domain, human, organizational and social intelligence
- The integration of agents, services, organizational and social computing with data mining for engineering agent mining symbionts [13]
- The integration of semantics, visualization, service and knowledge management for agent mining systems and applications
- Building trust, reputation, privacy and security for agent mining systems, and making sure the systems and results are sound and safe
- The development of an analytical methodology for agent retraining
- The development of a methodology for evaluating MAS efficiency, and the performance of the system as a whole as well as individual units.
- Measurements of gains achieved by the agent mining system for organization of decision
- Employing various distributed computing techniques to agent-based distributed data mining, such as the peer-to-peer model
- Ubiquitous computing, including ubiquitous knowledge discovery in ubiquitous environments for agent-mining integration and super-intelligent systems
- Privacy and security models for distributed agent systems. As they are traveling all over the system, it is necessary to guarantee privacy and security preservation, as in trust-based modeling
- The development of an agent infrastructure model that includes data mining as a key component, and vice versa
- The Web as a platform for agent mining, toward the use of future Web models, as Web 2.0 and 3.0, by considering the browser itself as a user agent which allows migration of problem domains and knowledge
- Successful systems, case studies and applications of agent mining technologies and systems for both professionals and common users.

1.10 Agent mining community development

The development of agent-mining interaction is evidenced by the following evolution process and burgeoning characteristics.

1.10.1 Fast progression in fostering the community

We draw a conclusion that agent-mining interaction and integration is emerging as a new member of the scientific family due to the following survey findings.

- An ever-increasing and respectable number of publications: an initial literature review has disclosed that there are over 250 conference and journal papers, books, proceedings, and technical reports published on topics associated with agent and mining interaction and integration. This uptrend is becoming increasingly clear from the incremental change over the past three years in the number of publications by major presses including Springer, IEEE and ISI press.
- An ever-increasing level and quality of papers: with the increase in publication numbers, publication quality has also been extremely improved. A typical trend is that there have been more and more journal papers and books/proceedings published after 2003. This is evidenced by increasing papers accepted by top-ranking conferences and journals such as AAMAS, KDD and ICDM in both communities.
- An ever-increasing number of professional activities: another typical indicator of whether a research topic is evolving into a new, separate area is the number and quality of professional activities, and the involvement of key research groups and researchers from both communities in these activities.
- An increasing transparent academic voice pursuing a separate area and a first-class citizen in the scientific family: this is evidenced by panel discussions in both ADMI workshop series and AIS-ADM series.

1.10.2 Research resources on agent-mining interaction

Agent mining communities are formed through emerging efforts in both AAMAS and KDD communities, as evidenced by continuous acceptance of papers on agent mining issues by prestigious conferences such as AAMAS, SIGKDD and ICDM. During the formation of agent mining communities, two special interest groups (SIG) have dedicated. AgentLink ALAD SIG was organized within the framework of FP5 European project KDNet Project, consolidating the data mining and agent communities. Another dedicated SIG is the Agent-Mining Interaction and Integration SIG [3], which actively links leading researchers in this emerging field, and or-

[3] AMII-SIG: www.agentmining.org

ganizes regular events for promoting and encouraging research networking in both agent and data mining communities. These task forces have spearheaded many other analogous initiatives including special sessions of the international conferences, dedicated international workshops, journal special issues, and more.

AMII-SIG now is one of the main and most active driving forces in promoting community development. It summarizes and shares information covering many issues, and has also led the annual workshop series on Agent and Data Mining Interaction (ADMI) since 2006, moving globally and attracting researchers from many different communities. It has also formed an international Steering Committee involving leading researchers from both agents and data mining communities.

The other biennial workshop series - Autonomous Intelligent Systems - Agents and Data Mining (AIS-ADM) has existed since 2005. Other events include special journal issues [16, 17], tutorials [42, 18], an edited book [9], and monographs [41, 19].

1.11 Conclusions

Agent and 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 open 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.

Acknowledgements This work is sponsored in part by Australian Research Council Grants (DP0988016, LP0989721, DP0773412, and LP0775041).

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. Baik,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://www-staff.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)
33. Letia, A., Craciun, F.; et. al.: First Experiments for Mining Sequential Patterns on Distributed Sites with Multi-agents. Intelligent Data Engineering and Automated Learning - IDEAL 2000: Data Mining, Financial Engineering, and Intelligent Agents, 19 Volume 1983 (2000)
34. Liu, J. and You, J.: Smart Shopper: An Agent Based Web Mining Approach to Internet Shopping. IEEE Transactions on Fuzzy Systems, Volume 11, Issue 2 (2003)
35. Mitkas, P.: Knowledge Discovery for Training Intelligent Agents: Methodology, Tools and Applications. Autonomous Intelligent Systems: Agents and Data Mining, Lecture Notes in Computer Science Volume 3505 (2005)
36. Lu, H.; Sterling, L. and Wyatt, A.: Knowledge Discovery in SportsFinder: An Agent to Extract Sports Results from the Web. PAKDD-99, Volume 1574 (1999)
37. Mohammadian, M.: Intelligent Agents for Data Mining and Information Retrieval, Idea Group Publishing (2004)
38. Ong, K., Zhang, Z., Ng, W. and Lim, E.: Agents and Stream Data Mining: A New Perspective. IEEE Intelligent Systems, Volume 20, Issue 3, 60 - 67
39. Rea, S.: Building Intelligent .NET Applications: Agents, Data Mining, Rule-Based Systems, and Speech Processing. Addison-Wesley Professional (2004)
40. Sian, S.: Extending Learning to Multiple Agents: Issues and a Model for Multi-Agent Machine Learning (MA-ML). In Proceedings of the European Workshop Sessions on Learning EWSL91 (Kodratroff, Y.) Springer-Verlag, (1991) 458-472.
41. Symeonidis, A., and Mitkas, P.: Agent Intelligence Through Data Mining, Springer (2006)
42. Symeonidis, A., and Mitkas, P.: Agent Intelligence Through Data Mining, *Tutorial with ECML/PKDD2006*.
43. Weiss, G.: A Multiagent Perspective of Parallel and Distributed Machine Learning. In *Proceedings of Agents'98*, 226-230, 1998.
44. Wooldridge, M.: An Introduction to Multi-Agent Systems, Wiley (2002)
45. Zhang, C.; Zhang, Z., and Cao, L.: Agents and Data Mining: Mutual Enhancement by Integration. Autonomous Intelligent Systems: Agents and Data Mining, Volume 3505 (2005)
46. Zhang, Z., and Zhang, C.: Agent-Based Hybrid Intelligent System for Data Mining. Agent-Based Hybrid Intelligent Systems, Volume 2938 (2004)
47. Zhong, N., Liu, J., and Sun, R.: Intelligent Agents and Data Mining for Cognitive Systems? Cognitive Systems Research Volume 5, Issue 3, (2004) 169-170