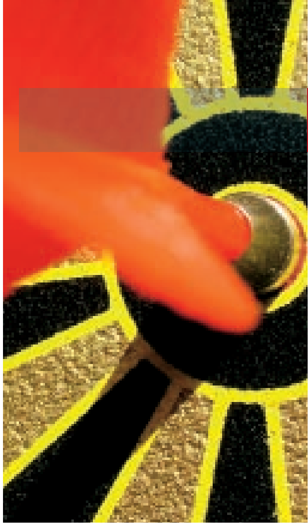

Contact Editor: **Brian Brannon,** bbrannon@computer.org

Agent Mining: The Synergy of Agents and Data Mining

Longbing Cao, *University of Technology, Sydney*
Vladimir Gorodetsky, *St. Petersburg Institute for Informatics and Automation, Russian Academy of Sciences*
Pericles A. Mitkas, *Aristotle University of Thessaloniki*

In the past twenty years, autonomous agents and multiagent systems (MASs, or agents) and data mining and knowledge discovery in databases (KDD, or data mining) have emerged as two of the most prominent, dynamic, and exciting research areas in information sciences. They've evolved with well-defined, yet distinct, aims and objectives to pursue methodologies and techniques to cope with the challenges of their respective areas.

Agents comprise a powerful technology for the analysis, design and implementation of autonomous intelligent systems that can handle distributed problem-solving, cooperation, coordination, communication, and organization in a multiplayer environment. It's also a promising computing paradigm for dealing with system complexities such as openness, distribution, human involvement, societal characteristics, and intelligence emergence. Agent research focuses on theoretical, methodological, technical, experimental, and practical issues and the means to handle system complexities. In addition to early research on individual, cognitive, and reactive agents, recent efforts have expanded to broader areas such as organizational and social problems in multiagent systems and society. Agents increasingly rely upon the foundational support of other disciplines including computer science, cognitive science, mathematics, social sciences, and human society.

For several decades, data mining (DM) and machine learning have retained their status among the topmost research and application areas of AI and intelligent information technology. DM aspires to respond to, arguably, the most difficult of the three main questions of artificial intelligence, which is, "Where does knowledge come from?" In the 1960s, experts understood that experimental data, both interpreted and uninterpreted, could provide rich sources of knowledge for intelligent systems. DM studies the process and techniques for analyzing massive data volumes to identify hidden, but interesting, patterns or relationships. With the ability to scrutinize ubiquitous and multimodal data sources, DM is successfully attacking emergent problems such as the discovery of patterns and knowledge in uncertain, high-frequency, organizational, or behavioral data, including data generated by multiagent systems. In addition to the mainstream focus on algorithm development and performance enhancement, DM has expanded to issues such as knowledge actionability, privacy processing, and human-centered mining. To this end, DM involves many disciplines, such as statistics, database technologies, machine learning, AI, pattern recognition, the Semantic Web, and behavior informatics.

Both fields face critical challenges that the other technology might alleviate. Typical problems in agents that could find satisfactory solutions in DM include multiagent learning, adaptation, evolution, and behavior analysis. For instance, knowledge extracted through DM could provide more stable, predictable, and controllable models for dispatching and planning, or it can assist in the self-organization and evolution of multiagent systems in acceptable directions. DM has also been widely used for behavior analysis, user modeling, and the development of services and recommendations in many domains (see "Research Resources on Agent Mining" sidebar). In general, data mining in available data logs can offer

1541-1672/09/$25.00 © 2009 IEEE
Published by the IEEE Computer Society

a much deeper behavioral understanding of MAS and provide more comprehensive, quantitative, and in-depth service solutions.

Agent-based data mining[1,2] somehow drives the emergence of the agent mining field. Agents can support and enhance the knowledge discovery process in many ways. For instance, agents can contribute to data selection, extraction, preprocessing, and integration, and they're an excellent choice for peer-to-peer parallel, distributed, or multi-source mining. Agents are also a good match for interactive mining, human-centered DM, service delivery, and customer service. Agent-based distributed DM[3–5] is a typical example of multiple agents collaborating with each other to mine for distributed data. In interactive mining, agents can bridge the gap between humans and software systems by acting as interfaces that can sense and affect human-mining needs. Mobile agents can also monitor and examine data and source changes and mine dynamic local patterns and then merge them into a combined pattern through coordination agents for conflict resolution. Taking the opposite route, software engineers can use data mining to extract knowledge models from large data sets. These models, in the form of decision trees or data-induced rules, can provide the logic for intelligent agents. Consider, for example, an enterprise resource planning (ERP) system that maintains a log of all actions and decisions a company takes. Using knowledge discovery techniques, the developer can identify, codify, and encapsulate the logic behind these actions into agents robust and trustworthy enough to replace human decision making. Figure 1 shows the two-way interaction between the two technologies.

Both agents and DM have to cope with issues such as domain knowledge, constraints, human roles and involvement, life-cycle and process management, and organizational and social factors. We can find enhanced solutions to these common problems in the integration of the two fields. Furthermore, many DM systems, like agents, are dynamic and need to accommodate online, runtime, and ad hoc requests. With the involvement of social intelligence and complexities, both areas have to worry about reliability, reputation, risk, security, and outcome actionability. Research in one area can stimulate, complement, and enhance research in the other.

In-depth analysis and comparison of both areas show not only an obvious need for interaction and integration but also an intrinsic linkage between agents and DM. Both areas share some similar or overlapping objectives related to the use of information and resources to deliver useful knowledge and intelligence from either an individual or a system perspective. There's also a significant overlap of theoretical underpinnings

Research Resources on Agent Mining

Agent mining communities are formed through emerging efforts in both the autonomous agents and multi-agent systems (AAMAS) and knowledge discovery and data mining (KDD) communities, as evidenced by continuous acceptance of papers on agent-mining issues by prestigious conferences such as AAMAS, SIGKDD and the IEEE International Conference on Data Mining (ICDM). A Special Interest Group on Agent-Mining Interaction and Integration (AMII-SIG; www.agentmining.org) was formed to share information 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. Another biennial workshop series, Autonomous Intelligent Systems—Agents and Data Mining (AIS-ADM), has existed since 2005. Other events include special issues in journals[1] (such as this one), tutorials, edited books,[2–4] and monographs.[5]

References

1. L. Cao et al., "Editor's Introduction: Interaction between Agents and Data Mining," *Int'l J. Intelligent Information and Database Systems*, vol. 2, no. 1, 2008, pp. 1–5.
2. V. Gorodetsky et al., eds., *Autonomous Intelligent Systems: Agents and Data Mining*, LNAI 3505, Springer, 2005.
3. V. Gorodetsky et al., eds., *Autonomous Intelligent Systems: Agents and Data Mining*, LNAI 4476, Springer, 2007.
4. L. Cao, ed., *Data Mining and Multiagent Integration*, Springer, 2009.
5. A. Symeonidis and P. Mitkas, *Agent Intelligence through Data Mining*, Springer, 2005.

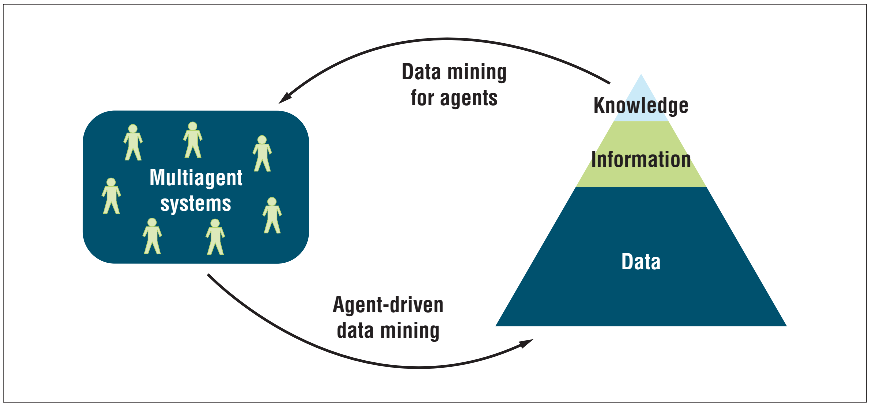

Figure 1. The synergy of agents and data mining. Agents can be used to mine large data collections efficiently. Knowledge models extracted via data mining can make agents more efficient.

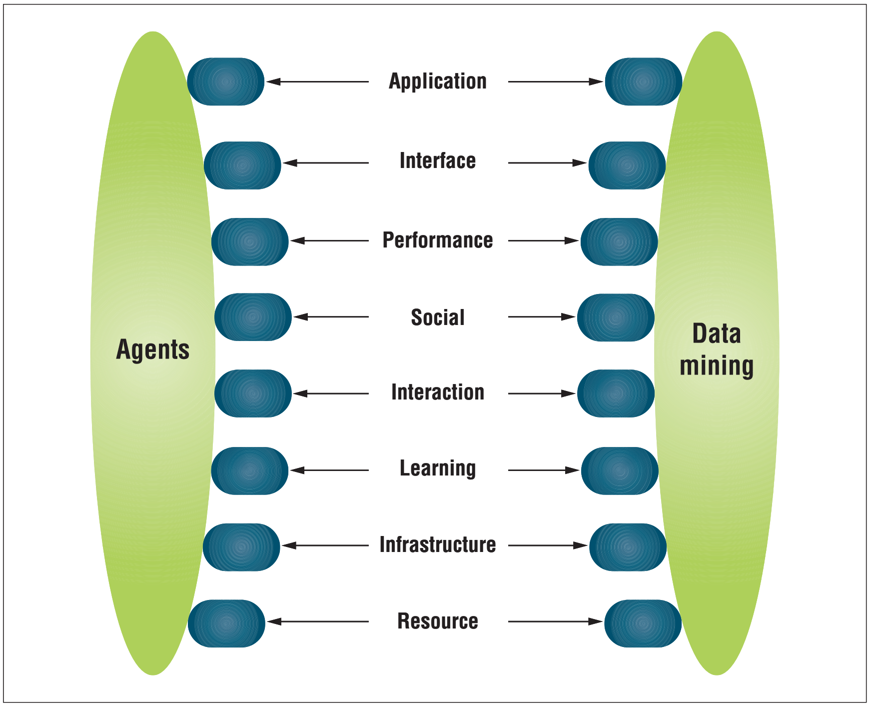

Figure 2. The agent-mining synergy may occur and can be analyzed in many dimensions.

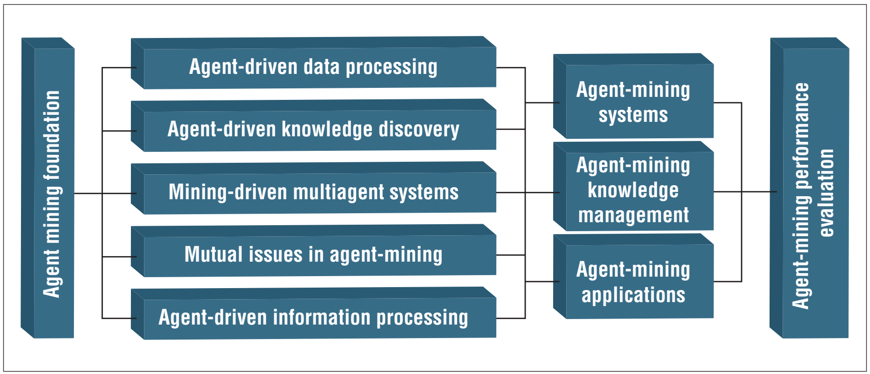

Figure 3. Agent-mining disciplinary framework.

and technical foundations. In addition, both areas are seeing increasing demands from business and other applications, such as personalized Web services, intrusion detection, peer-to-peer computing, and so on. These all involve organizational, environmental, and social factors and have to deal with domain factors and human involvement.

With the increasing need for interaction and integration between agents and DM, research in this new and promising area has elevated agent mining to the forefront of science and technology. Researchers have made great strides in agent mining from theoretical foundation-building to the development of technical means and tools. More and more applications are benefiting from the synergy of agents and DM.

A Disciplinary Framework

Agent mining refers to the methodologies, principles, techniques, and applications for the integration and interaction of agents and data mining, as well as the community that focuses on the study of the complementarity between these two technologies. As shown in Figure 2, we can analyze the agent mining synergy in different dimensions in that synergy can occur

- *in the resource dimension* at the data, information, and knowledge levels;
- *in the infrastructure dimension* in infrastructure, architecture, and process;
- *in the learning dimension* in learning methods, learning capabilities, and performance;
- *in the interaction dimension* in coordination, cooperation, negotiation, and communication;
- *in the social dimension* in social and organizational factors—for instance, in human roles;
- *in the performance dimension* in the performance enhancement of one end of the coupled system;
- *in the interface dimension* in the human-system interface, user modeling and interface design; and
- *in the application dimension* in applications and domain problems.

From these dimensions, many fundamental research issues and problems in agent mining emerge. Correspondingly, we can generate a high-level research map of agent mining as a disciplinary area. Figure 3 shows such a framework, which consists of the following research components:

- *Agent mining foundations* studies issues such as challenges and prospects, research maps and theoretical underpinnings, frameworks and so on.
- *Agent-driven data processing* studies issues associated with the extraction, collection, and management of data produced by the actions of agents in MAS. Agent-monitoring tools are usually required and must be developed and deployed in ways that won't interfere with the agents' operation. Tools for processing and que-

rying agent data are also essential.

- *Agent-driven knowledge discovery* focuses mainly on the problems related to the modeling, execution and optimization of the knowledge discovery process. The issue of automated data learning in dynamic or interactive environments is of paramount importance. Agents could prove more suitable in parallel and distributed architectures, such as the Web or the grid, where peer-to-peer learning and cooperation are desired. Agents are inherently better equipped for mining multiple, often heterogeneous, data sources.
- *Mining-driven multiagent systems* studies issues related to the proper and formal representation of knowledge models extracted from application data. Embedding data-induced inference engines into dummy agents requires an arsenal of development and verification tools that guarantee the soundness of the ensuing agent decisions. Agent training, adaptation, evolution, and learning are important issues, especially in self-organized and self-learning MASs. Agents must progressively build their reputations through risk- and trust-analysis mechanisms.
- *Agent-driven information processing* deals with issues such as information gathering and retrieval in an MAS, pattern analysis, and association studies, as well as multiagent domain intelligence involvement.
- *Mutual issues in agent mining* includes an array of issues that pertain to both technologies, such as human-system interaction, infrastructure and architecture problems, ontology and semantic issues, and so on.
- *Knowledge management* is essential for both agents and data mining, and it's the same for agent mining. This involves the representation, management, and use of ontologies, domain knowledge, human empirical knowledge, metadata, and metaknowledge in the agent-mining symbionts. Formal methods and tools are necessary for modeling, representing, and managing knowledge. These techniques should also cater to identifying and distributing knowledge, knowledge evolution in agents, and enabling knowledge use.
- *Agent-mining performance evaluation* researches methodologies, frameworks, tools, and test beds for evaluating the performance of agent mining, performance benchmarking, and metrics. Besides technical performance, which usually includes speed, accuracy, and statistical significance, business-oriented performance metrics such as cost, benefits, and risk are also important.
- *Agent-mining systems*. This component offers a system's perspective and studies techniques for the formation, modeling, design, and software engineering of agent-mining systems. Frameworks can be developed for extending existing applications into agent-mining systems or for tailoring such a system to a specific application. From our experience, the development of dedicated platforms for agent-mining systems is not a trivial task.
- *Agent-mining applications* refers to any real-world applications and domain problems that can be better handled by agent-mining technologies. Based on the needs of particular applications, any issues discussed in such topics can be engaged here. For instance, in some cases, we need to build an agent-mining simulation system to understand the working mechanism and potential optimization of a complex social network. In other cases, the enhancement of learning capability is the main issue, and we must use appropriate learning tools on demand.

> Agents are inherently better equipped for mining multiple, often heterogeneous, data sources.

In the remainder of this article, we differentiate between the two cases of agent-driven DM and DM-driven agents. In each case we outline the technological challenges that justify the need for each approach.

Agent-Driven Distributed Data Mining

This section discusses state-of-the-art agent-driven knowledge discovery with an emphasis on distributed DM. Agent-driven data mining has grown to become the most popular and prolific field of study since the inception of agent mining.

The Challenges of Distributed Data Mining

DM and machine learning form a mature field of AI supported by a variety of approaches, algorithms, and software tools. However, modern requirements for data mining and machine learning, inspired by emerging applications and information technologies and the peculiarities of data sources, continue to become increasingly more difficult. The critical features of data sources that determine 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 sources are of high frequency or density, mixing static and dynamic data and multiple data structures.

- Data integration and matching are difficult to conduct; it's impossible to store them in centralized storage, and it's infeasible to process them in a centralized manner.
- In some cases, multiple sources of data are stored in parallel storage systems.
- Local data sources can have restricted availability due to privacy and their commercial value, which in many cases can also prevent centralized processing, even in a collaborative mode.
- In many cases, distributed data spreading across global storage systems is often associated with time differences.
- The availability of data sources in a mobile environment depends on time.
- The infrastructure and architecture weaknesses of existing distributed data mining systems require more flexible, intelligent, and scalable support.

These and other peculiarities dictate the development of new approaches and technologies in data mining to identify patterns in distributed data. Multiagent technology dovetails nicely with the requirements of distributed data mining (DDM) and, more particularly, peer-to-peer (P2P) data mining.

The Need for Agent-Driven Distributed Data Mining

While analyzing the challenges associated with the practical implementation of distributed and P2P data mining, Matthias Klusch and his colleagues argue that agent technology is best able to cope with them in terms of autonomy, interaction, dynamic selection and gathering, scalability, multistrategy, and collaboration.[4] In addition to privacy, mobility, time constraints (in stream data, it's too late to extract and then mine), computational costs, and performance requirements, other reasons include the following:

- *Isolation of data sources*. Distributed and multiple data sources are often isolated from each other. For in-depth understanding of a business problem, it's essential to bring relevant data together through centralized integration or localized communication. Agent planning, collaboration, communication, and negotiation, as well as mobile agents can help.
- *Mobility of source data and computational devices*. Data and device mobility require the perception and action of data mining algorithms on a mobile basis. Mobile agents can adapt to mobility very well.
- *Interactive DDM*. Agents that provide proactive assistance are necessary to limit how much the user has to oversee 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 is to discover and select relevant sources (for instance, pursuing DM tasks where the availability of data sites and their content can change at any time). In these settings, DM agents could be deployed to adaptively select data sources according to given criteria such as the expected amount, type, and quality of data at the considered source, as well as the actual network and DM server load.
- *Time constraints on distributed data sources*. Time differences affect data distributed in different types of storage. Agents can be used to coordinate the process of data gathering.
- *Multistrategy DDM*. For some complex application settings, an appropriate combination of multiple data mining techniques can be more beneficial than applying just one. DM agents can learn which deliberative actions to choose on the basis of data retrieved from different sites and the mining tasks they pursue.
- *Collaborative DDM*. DM agents can operate independently on data they've gathered at local sites and then combine their respective models. Alternatively, they can agree to share knowledge as it's discovered to benefit from the additional options of other DM agents.
- *Privacy of source data*. Privacy issues restrict distributed local data from being extracted and directly integrated with other sources. A DM agent with authority to access and process the data locally can dispatch identified local patterns for engagement with findings from other sources.
- *Organizational constraint on distributed data sources*. In some organizations, business logic, processes, and workflows determine the order of data storage and access, which can make DDM more complex and delay the distribution of DDM algorithm agents among agents in different storage areas.

Web mining is a formidable and demanding representative of DDM that exhibits most of the challenges described above. Knowledge discovery in the Web takes place increasingly by the use of agents.

Agents that provide proactive assistance are necessary to limit how much the user has to oversee the data mining process.

Data Mining-Driven Agents

This section discusses DM-driven multiagent systems in which agents are empowered by knowledge models derived through data mining.

The Challenges of Data-Mining-Driven Agents

The astonishing rate at which current applications generate and collect data is difficult for even the most powerful of today's computer systems to handle. This windfall of information often requires another level of distillation to elicit knowledge hidden in voluminous data repositories. Data mining can extract knowledge nuggets that can constitute the building blocks of agent intelligence. Here, we loosely define intelligence 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 we can easily deduce or induce, 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 can't accommodate the dynamic transfer of DM results to the agents. To enable the incorporation of dynamic, complex, and reusable rules in multiagent applications, we must adopt a systematic approach.

DM techniques can filter existing application data—such as past transactions, decisions, data logs, agent actions, and so on—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 DM-driven agent systems, induction attempts to transform specific data and information into generalized knowledge models. The induction process produces new rules and correlations aimed at validating user hypotheses. Since induction is based on progressive generalizations of specific examples, it can 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. MAS applications use deduction 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.

> Data mining can extract knowledge nuggets that can constitute the building blocks of agent intelligence.

The coupling of the induction and deduction 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.

Knowledge Transfer from DM to Agents

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 agents already in operation in an agent-based application. We consider three distinct cases, which correspond to three types of knowledge extracted as well as different data sources and mining techniques:

- *Case 1*—knowledge extracted by performing DM on historical datasets that record the business logic (at a macroscopic level) of a certain application;
- *Case 2*—knowledge extracted by performing DM on log files that record 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 the ability to dynamically embed the extracted knowledge models into the agents and repeat this process as many times as necessary. The developer can follow standard agent-oriented software engineering processes to specify the application ontology, the agent behaviors and agent types, the communication protocol between the agents, and their interactions.

Andreas Symeonidis and Pericles Mitkas describe a number of agent-based applications that cover all three cases of knowledge diffusion.[6] 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 that tracks user actions in corporate sites and suggests other sites of interest. A large variety of Web services and intranet applications can benefit from a similar approach.

Finally, Case 3 encompasses solutions for ecosystem modeling and for Web crawling with clusters of synergetic crawler agents.

Agent Academy is an open-source framework and integrated development environment for creating software agents and multiagent systems and for augmenting agent intelligence through data mining.[7] Agent Academy has been implemented on the JADE and WEKA APIs. Its second version is now available at Sourceforge (http://sourceforge.net/projects/agentacademy).

Advantages of Data-Mining-Driven Agents

DM-driven multiagent systems present attractive features that can lead to more intelligent systems, including the following:

- 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 agents' learning and knowledge processing capability 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 they make planning and reasoning decisions.
- DM can enhance the agent capability of handling uncertainty via historical event analysis, dynamic mining, and active learning. By mining agent behavioral data, it's possible to reach a balance between agent autonomy and supervised evolution. Thus, 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 provide agents the capability of learning, discovery, probing, and searching.
- Real-world databases often contain missing, erroneous data, or outliers. Clustering can assimilate noisy logs so they become part of a greater group, thus smoothing down differences while detecting and rejecting outliers. Classification can validate ambiguous data records and estimate missing data records. Rule-based systems can't handle such data efficiently without increasing their knowledge-base and therefore their maintenance cost.

> By mining agent behavioral data, it's possible to reach a balance between agent autonomy and supervised evolution.

Our approach favors the combination of inductive and deductive reasoning models. Agents can deploy deductive reasoning models to ensure system soundness using preprocessed data from inductive agents. This satisfies the application domains' dynamic nature while the set of deductive results (knowledge-bases of deductive agents) becomes more compressed and robust.

Although it is not always easy to define the patterns and rules generated through data mining as being sound, a software engineer can deploy metrics to evaluate the algorithms' performance. Such metrics may include total mean square error (clustering), support-confidence (association rules), classifier accuracy (classification), and classifier evaluation (genetic algorithms), among others.

A developer usually introduces DM tools to enterprises as off-the-shelf components. Human-experts use these tools to examine their corporate or environmental databases to develop strategies and make decisions. This 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, one can diffuse useful recommendations much faster and apply parallelism to unrelated tasks, pushing system performance even higher.

Mutual Issues in Agents and Data Mining

Many mutual enhancement and integration issues exist in both agents and data mining. A typical issue involves human intelligence and human roles. Although both communities recognize the importance of human involvement and human intelligence in problem solving and solution development, it's challenging to effectively and dynamically include human roles in problem-solving systems. Issues arise from aspects such as understanding and simulating human empirical intelligence and experiences critical to problem solving, acquisition and representation of human qualitative intelligence in agent-mining systems, and the interaction and interfaces between humans and systems catering to human intelligence and roles.

Organizational, environmental, and social factors constitute important elements of complex problems in agents and data mining. These include comprehensive factors such as business processes, workflows, and business rules, as well as human roles relevant to problem solving and organizational and social factors including organizational rules, protocols, and norms. For instance, while concepts such as organizational rules, protocols, and norms have been fed into agent organizations, they're also important for DM systems in converting patterns into operable deliverables that business people can smoothly take over and integrate into business systems.

There are often gaps between technical outcomes and business expectations in developing workable agents and DM algorithms and systems due to the inconsistent and incomplete evaluation methodologies. As a result, business people are often not interested in the resulting deliverables which don't work for problem solving. An ideal scenario is to generate algorithms and systems that care about technical and business concerns from objective and subjective perspectives.

These case studies show that it's essential to study common issues for the benefits of a particular field. In fact, studies can also activate the emergence of agent-mining symbionts. For instance, the two technologies can share the modeling and representation of domain knowledge and knowledge management. This can serve as an intrinsic working mechanism for an agent-mining symbiont that has the capability of involving domain knowledge in agent-human interaction, DM algorithm modeling, and knowledge management for DM agents and agent-based systems.

Additional mutual issues include architecture and infrastructure problems, handling of constraints and business requirements, production of metadata and metaknowledge, and the more mundane business factors of security, privacy and trust.

To facilitate the study of common enhancement issues, researchers might need to develop diversified and cross-disciplinary methodologies. Useful bodies of knowledge could include cognitive science, human-machine interaction, and interface design, knowledge engineering and management, ontological engineering, and so on.

> As many references show, the idea of agent mining is actually driven by broad and increasing applications.

Application Areas

As many references show, the idea 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 intelligent information processing. Application domains include:

- e-commerce, including electronic auctions, negotiations, and trading,
- business intelligence, including supply chain and customer relationship management,
- DM in applications that need decision support and quick action, such as financial markets, healthcare, and security,
- resource allocation in parallel and distributed environments, such as grid computing, network services, parallel computing,
- mining Web 1.0, 2.0 and the Semantic Web,
- P2P DM, and so on.

These are just a few of the areas where researchers have suggested, explored or demonstrated that agent mining can better handle problems as opposed to using unilateral technology.

Open Issues

Open issues and questions still remain in this synergy. Research and exploitation are on the way to ensure better, tighter, and more organized system integration. Open issues that still require resolution include:

- Foundations, including methodologies, formal tools, and processes for supporting the integration of agents and DM from multiple dimensions as outlined in Figure 2.
- Formulation of formal methodologies, languages, and notations for agent mining software engineering.
- 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, including individual unit performance and the performance of the system as a whole.
- Employing various distributed computing techniques to agent-based distributed data mining, such as the peer-to-peer model.
- Measurement of gains achieved by the agent mining system for decision making, and so on.

Data mining is a mature, widely-accepted field while multiagent systems are gaining ground steadily. Agent mining seems well poised to proceed from the stage of prototypes to the development of successful systems, case studies, and applications for both professionals and common users. The remaining open issues in the integration of the two technologies represent a fertile ground for research.

This article presents an overall picture of agent mining, a young scientific field that advocates and studies the integration of agent technology and data mining. The analysis and synthesis of challenges, strengths and weaknesses in each field, and the mutual and extra benefits through synergy, clearly indicate the need for and the promising potential of agent mining for the mutual enhancement of both fields and for the creation of super-intelligent systems. Even though many researchers have been committed to both areas, more efforts are required to develop techniques, systems, and case studies from foundational, technological, and practical perspectives.

Acknowledgments

L.B. Cao wishes to acknowledge the support of Australian Research Discovery grants DP0773412 and DP0667060 and Linkage grant LP0775041.

References

1. W. Davies and P. Edwards, "Agent-Based Knowledge Discovery," *Working Notes of the AAAI Spring Symp. Information Gathering from Heterogeneous, Distributed Environments*, AAAI Press, 1995, pp. 34–37.
2. H. Kargupta, I. Hamzaoblu, and B. Stafford, "Scalable Distributed Data Mining Using an Agent-Based Architecture," *Proc. 3rd Int'l Conf. Knowledge Discovery and Data Mining* (KDD 97), 1997, pp. 211–214.
3. G. Weiss, "A Multiagent Perspective of Parallel and Distributed Machine Learning," *Proc. 2nd Int'l Conf. Autonomous Agents* (Agents 98), ACM Press, 1998, pp. 226–230.
4. M. Klusch, S. Lodi, and G. Moro, "Agent-Based Distributed Data Mining: The KDEC Scheme," *AgentLink*, LNCS 2586, Springer, 2003, pp. 104–122.
5. V. Gorodetsky, O. Karsaev, and V. Samoilov, "Infrastructural Issues for Agent-Based Distributed Learning," *IEEE/WIC/ACM Int'l Conf. Web Intelligence and Intelligent Agent Technology—Workshops* (IADM 06), 2006, pp. 3–6.
6. A. Symeonidis and P. Mitkas, *Agent Intelligence through Data Mining*, Springer, 2005.
7. P.A. Mitkas et al., "A Framework for Constructing Multiagent Applications and Training Intelligent Agents," *Agent-Oriented Software Engineering IV*, LNCS 2935, Springer, 2003, pp. 96–109.