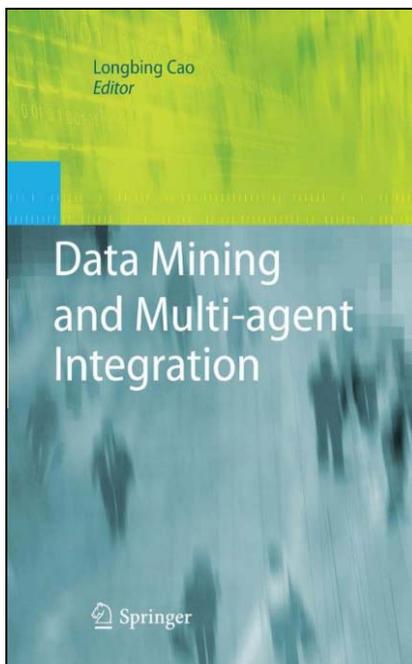


Data Mining and Multi-agent Integration

BY LONGBING CAO (EDITOR) – ISBN: 978-1-44190-521-5



REVIEWED BY ANDREAS SYMEONIDIS

Nowadays, both Agent Technology (in the form of individual agents, multi-agent systems, and societies) and Data Mining technologies have reached an acceptable level of maturity, and, each one alone, has its own scope and applicability. Naturally enough, a fruitful synergy of the two technologies has already been proposed, that would combine the benefits of both worlds and would offer computer scientists with new tools in their effort to build more sophisticated software systems. This integration is at least bidirectional: on the one hand, the *Agent-driven Data Mining* approach identifies that, since Data Mining comprises a number of discrete, nevertheless dependent tasks, agents can be employed in order to regulate, control, and organize the, potentially distributed activities involved in the knowledge discovery process. On the other hand, the *Data Mining-driven Agents* approach argues that knowledge hidden in voluminous data repositories, social networks and

web transactions, can be extracted by data mining and provide the inference mechanisms or simply the behavior of agents and multi-agent systems. In other words, the discovered knowledge nuggets may constitute the building blocks of agent intelligence.

However, while the pieces are already there, the puzzle is far from complete. Coupling the two technologies does not come seamlessly, since the inductive nature of data mining imposes logic limitations and hinders the application of the extracted knowledge on deductive systems, such as multi-agent systems. One should take all the relevant limitations and considerations into account, in order to provide a pathway for employing data mining techniques in order to augment agent intelligence.

“Data Mining and MultiAgent Integration”, edited by Longbing Cao, one of the experts in the field and founder of the AMII Special Interest Group, focuses on exactly this synergy between Agent Systems and Data Mining. *Agent Mining*, as defined in the latest bibliography, is expected to create innovative interaction and integration tools and services, and unify results under one new technology.

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

state-of-the-art and challenges through a number of research cases.

Part I is organized in three chapters. Chapter 1, written by the editor, serves as a synopsis of the content to follow. It pinpoints the main driving forces of the new technology, and summarizes the disciplinary framework, case studies, trends and directions currently in the field. Based on studying issues related to agent-driven data mining, data mining-driven agents and their interdependence, Chapter 1 acknowledges the potential of the new technology and depicts the theoretical and practical issues that the integration has. Chapter 2 follows a bottom-up approach. Through two pilot case demonstrators, a MAS for supporting manual annotation for DNA function prediction and a MAS to assist in digital forensics, an effort is made to clarify the benefits of Data mining-driven agents. Finally, Chapter 3, following again a bottom-up approach, performs a thorough survey and provides evidence on the exploitation of agents in distributed data mining, in terms of significance, system architectures, and research trends.

Having sketched the bigger picture in Part I, Part II provides the reader with design and implementation details in a variety of problems solved through the use of data mining-driven agents. Chapter 4 presents an agent system for web session clustering, based on swarm intelligence. Chapter 5 focuses on improving agent intelligence through discovering temporal agent behavior patterns, while Chapter 6 employs Web usage and Web structure mining in order to analyze user interaction habits and predict user behavior. In the same context. Chapter 7 presents a distributed recommender system and the methodology employed for sharing and

generating improved recommendations. Chapter 8 integrates a multi-class supervised classification algorithm with agent technology in the domain of network intrusion detection, where a multi-agent design methodology is coupled with a highly accurate, fast, and lightweight *PCC Classifier* and *CRSPM* schemes. Chapter 9 proposes genetic algorithms for data extraction based on the evolution and grammatical composition of regular expressions, while Chapter 10 employs a weight-driven network module in order to increase projection of knowledge nodes in a system, enrich their repositories and stimulate the corresponding user communities. Chapter 11 defines the notion of Goal mining and utilizes it in order to extract knowledge on user goals, residing in common query logs. Finally, Chapter 12 concludes the collection of manuscripts on data mining driven agents, discussing an agent-based diagnostic workbench equipped with classification capabilities, in order to support real medical diagnosis.

Part III addresses the complementary to Part II approach. First, in Chapter 13, EMADS is presented a framework that extends current work in data mining frameworks and employs different types of cooperating agents in order to perform complex classification and association rule extraction tasks. Next, Chapter 14 proposes a multi-agent system for dealing with online hierarchical clustering of streaming data, while Chapter 15 proposes two models for Agents-driven clustering of large datasets: a divide-and-conquer method and a data-dependent method. Chapter 16 introduces MACC, a multi-ant colony and multi-objective clustering algorithm that handles distributed data, through the assignment of specific objectives to different colonies and the synthesis of all results. Chapter 17 proposes an interactive environment for psychometrics diagnostics, where agents monitor user actions and perform data mining on them, in order to discover potentially interesting information. Chapter 18, following a completely different approach, implements a two-level agent system that performs association rule

mining and frequency based mining on log files, in order to discover firewall policy rules, subsequently employed to detect intra- and inter- firewall anomalies. Chapter 19 employs simple data mining and statistical analysis on a heterogeneous data grid and proposes a game theory-based multi-agent model for competitive knowledge extraction, hierarchical knowledge mining, and Dempster-Shafer result combination. Chapter 20 discusses a normative multi-agent enriched data mining architecture and ontology framework to support citizens in accessing services provided by public authorities. Chapter 21 works on the combination of static and dynamic agent societies assigned with the task of identifying (though classification) groups of users with common interests. Finally, Chapter 22, the last Chapter of Part III and the book, describes an agent-based video contents identification scheme using a watermark based filtering technique, aiming to prevent a user from uploading illegal video content into a user defined web storage.

Overall, reading this book is a pleasant surprise. The editor, having satisfied good quality contributions from the authors, has succeeded in producing a book that may serve as the basis for further probing. The objectives and expected outcome of reading the book become clear from the very beginning, the structure is concise and the pilot cases provided with respect to the two established lines of work in agent mining are representative. Having read "*Data Mining and MultiAgent Integration*", the user is triggered to explore the issues related to the coupling of the two technologies, deciding to follow any agent mining path, either one already established, or a completely new one. It is with interest that we expect the L. Cao et al monograph, "*Agents and Data Mining: Interaction and Integration*".

ABOUT THE REVIEWER:

ANDREAS SYMEONIDIS:
Department of Electrical and Computer Engineering, Aristotle University of Thessaloniki, Greece. Contact him at: asymeon@eng.auth.gr and see <http://users.auth.gr/symeonid> for more information.