Medical Informatics Europe 2002
G Surján et al. (eds)
IOS Press 2002

Leveraging Intelligent Agents for Knowledge Discovery From Heterogeneous Healthcare Data Repositories

Syed Zahid Hassan ZAIDI[1], Syed Sibte Raza ABIDI[2] & Selvakumar MANICKAM[1]

[1]*Health Informatics Research Group, School of Computer Sciences*
Universiti Sains Malaysia, Penang, MALAYSIA
[2]*Faculty of Computer Science, Dalhousie University, Halifax B3H 1W5, CANADA*

Abstract. This paper presents a case for an intelligent agent based framework for knowledge discovery in a distributed healthcare environment comprising multiple heterogeneous healthcare data repositories. Data-mediated knowledge discovery, especially from multiple heterogeneous data resources, is a tedious process and imposes significant operational constraints on end-users. We demonstrate that autonomous, reactive and proactive intelligent agents provide an opportunity to generate end-user oriented, packaged, value-added decision-support/strategic planning services for healthcare professionals, manages and policy makers, without the need for a priori technical knowledge. Since effective healthcare is grounded in good communication, experience sharing, continuous learning and proactive actions, we use intelligent agents to implement an *Agent based Data Mining Infostructure* that provides a suite of healthcare-oriented decision-support/strategic planning services.

1. Introduction

Over the past several years, there has been an explosion in the amount of healthcare data generated and subsequently collected in biomedical/healthcare data warehouses. This abundance of healthcare data has resulted in a large number of concerted efforts to inductively discover 'useful' knowledge from the collected data, and indeed interesting results have been reported by health informatics researchers [1] [2]. However, despite the noted efficacy of knowledge discovery methods—known as *Data Mining (DM) algorithms*—the challenge facing data managers today is about: (i) *data accessibility*—i.e. how to fully access and seamlessly utilize heterogeneous healthcare data that may exist in varying modalities, semantics, data formats and across multiple institutions in a healthcare enterprise; and (ii) *data usability and impact*—i.e. the use of 'appropriate' data mining algorithms with the 'right' data to discover value-added 'action-oriented' knowledge in terms of data-mediated decision-support/strategic-planning services.

We argue that for effective healthcare DM, healthcare data repositories—the backend data resource to any data mining system—need not be envisaged as independent entities. Rather, with the existence of multiple heterogeneous data repositories in a healthcare enterprise we need to establish a distributed data community, such that any DM effort draws upon the 'holistic' data available within the entire healthcare enterprise. When adopting this point-of-view, a set of data access and mining issues can be addressed using the intelligent agent paradigm, which can then be embraced for its analysis and implementation [3]. The aim of this paper is to introduce the intelligent agent paradigm for distributed DM from heterogeneous healthcare repositories, and then to describe a prototype agent-based architecture for healthcare DM. We present an *Agent-based DM Info-structure (ADMI)*, that deploys a suite of DM algorithms coupled with intelligent agents to facilitate data access, DM query specification, DM algorithm selection and DM result visualisation—i.e. automated generation of data-mediated decision-support/strategic-planning services.

2. Intelligent Agent Paradigm

The underlying concept for defining agents is the intentional system, developed by Dennett [4], which identifies an entity with a behavior predictable by attributing belief, desires, and rational insight to it. Although there is no widely agreed upon definition of agents, the following properties are commonly accepted:

- *Autonomy*: an agent has control over its own actions, and maintains an internal state that enables it to function without user supervision or support;
- *Sociality*: an agent interacts with others by means of an agent communication language (ACL);
- *Reactivity*: an agent models the environment in which it operates and acts as a consequence of changes in the environment;
- *Proactivity*: an agent acts to pursue its design aims, thus showing an opportunistic behavior. An Agent may embody intelligent techniques to exhibit intelligent behaviour.
- *Adaptivity*: Over time, an agent can adapt to a behavior that suits the preferences of an individual user or the dynamics of the operating environment.
- *Continuity*: Unlike programs that are selectively invoked to perform a particular software task, an agent can exist in an environment in a dormant state and begin operation 'as and when needed'.

With these functional characteristics agents provide useful metaphors for describing artificial systems, such as the following:

- *open systems*, which are dynamically changing because they are based on heterogeneous components, appearing, disappearing, and changing behavior;
- *complex systems*, where the agent paradigm provides a way of abstracting and analyzing them;
- *systems with distributed data, control, or resources*, where solutions are given for their design and implementation.

In the last few years, a number of agent-based applications have emerged in the biomedical/healthcare field , a prominent one is the GUARDIAN system [5], devoted to patient monitoring in a surgical intensive-care unit. In such a system, support is provided for collaboration among specialists, each expert in a specific domain, but committed to share data and knowledge with the others, and nurses that continuously monitor the patient under the physician's control.

3, Intelligent Agents in a Data Mining Framework

The autonomous discovery of healthcare knowledge from distributed biomedical/healthcare data repositories seems to be a good candidate for the application of the intelligent agent paradigm [6]. We argue so because, the a typical inductive knowledge discovery framework involves (a) complex systems with heterogeneous DM components, (b) managing distributed and heterogeneous data repositories, (c) servicing semanticallydiverse queries from different types of users, and (d) synthesizing multiple DM outcomes to formulate pro-active, 'action-oriented' decision-support/strategic-planning services targeted for a variety of users, ranging from healthcare practitioners to administrators [7].

We identify an scenario in order to identify some of the relevant issues pertaining to the functional incorporation of intelligent agents in a DM framework. The scenario pictures a regional healthcare manager/administrator dealing with an infectious-disease epidemic and needs to (a) forecast the possible geographical impact of the epidemic; (b) analyse hospital admissions due to the epidemic; and (c) plan an effective antibiotic regime (w.r.t. to the sensitivity of different antibiotics to the infectious disease) to treat the infectious bacteria. Considering that a region comprises multiple hospitals, each with its own data repository that is remotely accessible via a private network, the tasks anticipated are as follows:

- Problem analysis and specification, which guides the choice of 'appropriate' DM service packages suited for individual tasks. Note that the dynamics of the DM algorithm impacts data preparation.
- Establishing a communication channel to enable remote access to the data repositories of multiple hospitals.
- Collection of 'relevant' data to complete each individual task need to be first identified and subsequently retrieved from the respective data repositories.
- Synthesis of heterogeneous data originating from multiple data repositories.
- Preparation of the data according to the specification of the DM service packages.
- Execution of the DM algorithm
- Generation of DM report for the end-user.

In this scenario, it may be noted that the tasks outlined above correspond well with the functional capabilities of intelligent agents, hence there is a case for a agent-based solution for healthcare-related DM (illustrated in Figure 1).

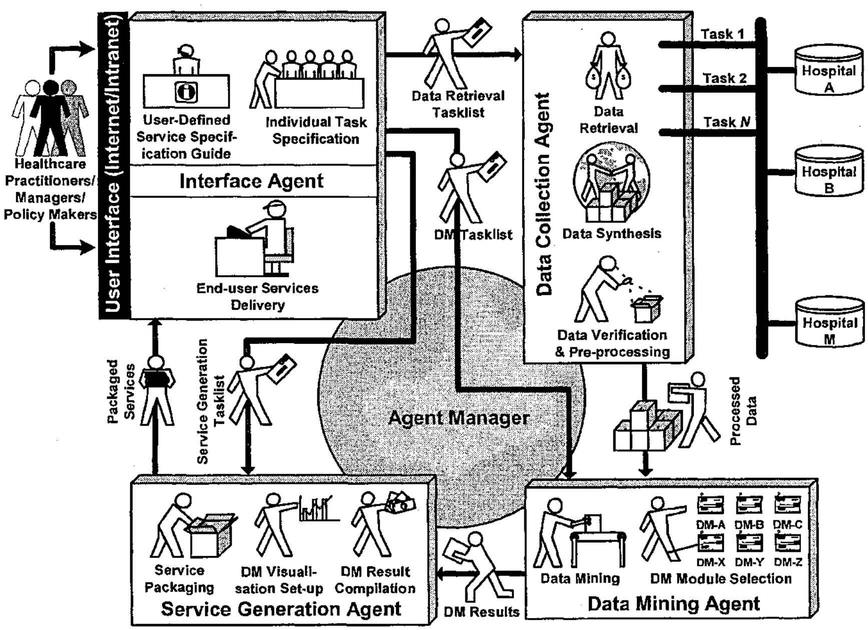

Figure 1: A sketch of an agent community—i.e. ADMI—which deals with the DM task given in scenario. Illutrated is the intrinsic behavior and inter-agent communication of the inherent intelligent agents. The shaded area represents the *Agent Manager*—a meta-agent responsible for coordinating all the agents.

4. Agent-Based Data Mining Info-Structure: An Overview

We present the architectural and functional overview of our proposed multi *Agent-Based Data Mining Info-Structure (ADMI)* for the generation of data-mediated diagnostic-support and strategic services. ADMI features a federation of intelligent agents, each one providing a particular service to either other intelligent agents or end-users. The agent-federation is designed to service four functional components—(i) end-user interface; (ii) remote data access network; (iii) data mining engine; and (iv) diagnostic-support and strategic services—and hence comprises of four autonomous intelligent agents which are described below.

4.1. Interface Agent

The *Interface Agent* (IA) manages a web-based graphical interface for the specification of a data-mediated service by the user. More specifically, the IA takes as input a set of service goals and autonomously translates them to three follow-up task specifications—(i) the DM tasks that need to be performed, the data that need to be retrieved and (iii) the results (i.e. the service) presentation style. This is achieved via the following modules:

- *Service Specification Module* comprises a list of pre-designed data-mediated services. Each service script guides the user to specify the parameters of the problem.
- *Task Definition Module* defines the follow-up tasks based on the service specification.
- *Communication Module* ensures communication with other agents mainly for assigning the processing tasks to the respective agents.

4.2. Data Collection Agent

The core functionality of the *Data Collection Agent* (DCA) is to facilitate the on-demand retrieval of 'relevant' data from the multiple healthcare data repositories. DCA is activated by IA which passes a task-specification to DCA. In an autonomous manner, DCA performs the following tasks: (a) establishing protocols for remote data access; (b) data selection and retrieval; (c) and data synthesis. The entire operational functionality of DCA is realized via the following inherent modules:

- *Data Collection Module* contains routines for translating a TA request to database queries wrt to the data model of the available data repositories.
- *Aggregation Module* defines constraints for heterogeneous data aggregation.
- *Pre-processing Module* processes the raw data to processing quality data via a variety of pre-processing techniques.
- *Communication Module* is responsible for communication with other agents for co-ordination purposes.

4.3. Data Mining Agent

The *Data Mining Agent* (DMA), as its name suggests, is responsible for co-ordinating the entire DM activities. It receives inputs from the IA (the DM task specification) and DCA (the pre-processed data). The DM tasks are classified into six classes, where each class may represent multiple DM algorithms. Each DM algorithm is implemented as a DM module. Given a task specification from the IA, the DMA autonomously selects the most appropriate DM module and coordinates all its processing requirements—i.e. data, constraints, protocols and so on. Technically speaking, each DM module comprises a script written in a high level declarative data mining language (comparable to SQL) that specifies the following characteristics: (a) classification of the DM task such as prediction, classification etc; (b) the DM algorithm; (c) attributes of data to be used, together with their significance values and relational information; (d) the possible output formats, such as reports, graphs, charts, tables, etc; and (e) the protocols and constraints imposed on the agent activities. Additionally, the DMA communicates with other agents for task execution, monitoring and results decomposition.

4.4. Services Generation Agent

The *Services Generation Agent* (SGA) processes the DM results produced by the DMA to generate decision-support/strategic services as per the user's request. The main activities of the SGA is to (a) collate all the results from the different DM modules; (b) set-up the result visualisation algorithms so as to produce different perspectives of the DM results; and (c) package the DM results as a turn-key decision-support/strategic service.

5. Conclusions

Healthcare data can be analyzed from different perspectives, each perspective imparting a different kind of knowledge—for instance hospital admissions data can be used to both analyze hospital admissions and also to forecast future hospital admission. In our opinion, multiple usage of healthcare data can be supported by the collaborative use of multiple, task-specific agent communities. This implies the need to take into account, and leverage as much as possible, the results of multiple intercommunicating agents, each working on some aspect of the data.

In our work, we have managed to leverage intelligent agents for generating (healthcare) data-mediated decision-support/strategic-planning services targeted for healthcare professionals and managers. At present, we plan to provide the following decision-support/strategic-planning services: *Analysing trends in hospital admission, Analysing treatment pattern, Analysing outcomes of treatment, Analysing cost-effectiveness of health care, Planning out-of hospital (ambulatory) care, Forecasting 'new disease' and strategising appropriate preventive measures, Forecasting complications of treatment, Forecasting the spread of infectious diseases.*

Although at a preliminary stage, the ADMI prototype developed thus far allows us to explore the ways functional collaboration may occur using agents to realise a distributed DM info-structure. The fact that all agents communicate through a common language, allows for an open-ended architecture whereby new DM services and DM algorithms can be incrementally added. Also, by incorporating the data retrieval logic within a number of individual DCA we have ensured the future inclusion of other healthcare institutions for data sharing. In conclusion, we believe that the emerging intelligent agent-based framework applied to the healthcare domain can provide interesting opportunities to operationalize the volumes of healthcare data routinely collected within numerous healthcare enterprises.

References

[1] Babic, Knowledge Discovery for Advanced Clinical data Management and Analysis, *P. Kokok et al. (Eds) Medical Informatics Europe'99*, Ljubljana, Amsterdam:IOS Press, 1999.

[2] S. S. R. Abidi, K. M. Hoe and A. Goh, Analyzing Data Clusters: A Rough Set Approach to Extract Cluster Defining Symbolic Rules. *Fisher, Hand, Hoffman, Adams (Eds.) Lecture Notes in Computer Science: Advances in Intelligent Data Analysis, 4th Intl. Symp of Intelligent Data Analysis IDA'01*, Berlin:Springer Verlag, 2001

[3] N. R. Jennings and M. J. Wooldridge, *Agent Technology Foundations, Applications, and Markets,* Berlin: Springer-Verlag, 1998.
[4] D. C. Dennett, *The Intentional Stance.* Cambridge, MA: MIT Press, 1987.
[5] Hayes-Roth and J. E. Larsson, A domain-specific software architecture for a class of intelligent patient monitoring systems, *J. Experimental Theoretical Artificial Intelligence*, vol. 8, pp. 149–171, 1996.
[6] H. Kargupta, I. Hamzaodlu, and B. Stafford, *Sclable, Distributed Data Mining Using an Agent Based Architecture,* D.Heckerman, et al (Eds) Proc. of 3rd Int.Conf. on knowledge discovery and Data Mining, AAAI press, pp.211-214, 1997.
[7] S. S. R. Abidi, Applying Knowledge Discovery in Healthcare: An Info-Structure for Delivering 'Knowledge-Driven' Strategic Services, *P. Kokok et al. (Eds) Medical Informatics Europe'99*, Ljubljana, Amsterdam:IOS Press, pp. 453-456, 1999.