In 15th IEEE Symposium on Computer Based Medical Systems (CBMS'2002), Maribor (Slovenia).

Distributed Data Mining From Heterogeneous Healthcare Data Repositories: Towards an Intelligent Agent-Based Framework

Syed Zahid Hassan Zaidi
Health Informatics Research Group,
School of Computer Sciences
Universiti Sains Malaysia, Penang, Malaysia

Syed Sibte Raza Abidi
Faculty of Computer Science,
Dalhousie University, Halifax B3H 1W5, Canada
Email: sraza@cs.dal.ca

Selvakumar Manickam
Health Informatics Research Group,
School of Computer Sciences
Universiti Sains Malaysia, Penang, Malaysia

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 and managers. We propose the use intelligent agents to implement a distributed Agent based Data Mining Infostructure that provides a suite of healthcare-oriented decision-support/strategic planning services.

1. Introduction

The proliferation 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 using a variety of *Data Mining (DM) algorithms* [1] [2].

We argue that 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 view, a set of data access and mining issues can be addressed using the well-known software agent technology [3]. The aim of this paper is to propose a prototype agent-based architecture to effectuate distributed DM activities applied to heterogeneous healthcare repositories. 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 Agents for Mining Healthcare Data

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 [4]. We identify a 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.
- Establishing a communication channel to enable remote access to the data repositories of multiple hospitals. Technically this involves the exchange of messages between agents over a network using a lower-level protocol such as SMTP, TCP/IP, or HTTP.
- 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 outlined tasks correspond well with the functional capabilities of intelligent agents, hence there is a case for an agent-based solution for healthcare-related DM activities (illustrated in Figure 1).

3. An Agent-Based Data Mining Info-Structure

Our proposed multi *Agent-Based Data Mining Info-Structure (ADMI)*, responsible for the generation of data-mediated diagnostic-support and strategic services, takes advantage of a multi-agent architecture which features the amalgamation of various types of intelligent agents, each responsible for an independent task. Such an 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. A brief overview of the constituent agents and their functionalities is presented below:

3.1. Interface Agent

The *Interface Agent* (IA) has been designed to collect user-specification for a DM service via a web-based interface. 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, (ii) the data that need to be retrieved and (iii) the results (i.e. the service) presentation style. This functionality is achieved via the following modules:

Service Specification Module comprises a library of pre-designed data-mediated services in the form of scripts. Each script assists the user to (a) correctly specify the various parameters of the problem; and (b) to choose the appropriate DM technique for

the problem at hand. The services can be defined by using *meta-pattern guided miner technique.* Which takes the user-specified services as a meta-rule form such as *"P (x, y)> q(y, z)—R (x, z)"* as a pattern to confine search for desired rules which are appropriate for different DM tasks.

Task Definition Module defines and delegates the tasks to the available agents.

Communication Module provides the mean for agents to exchange process-related messages.

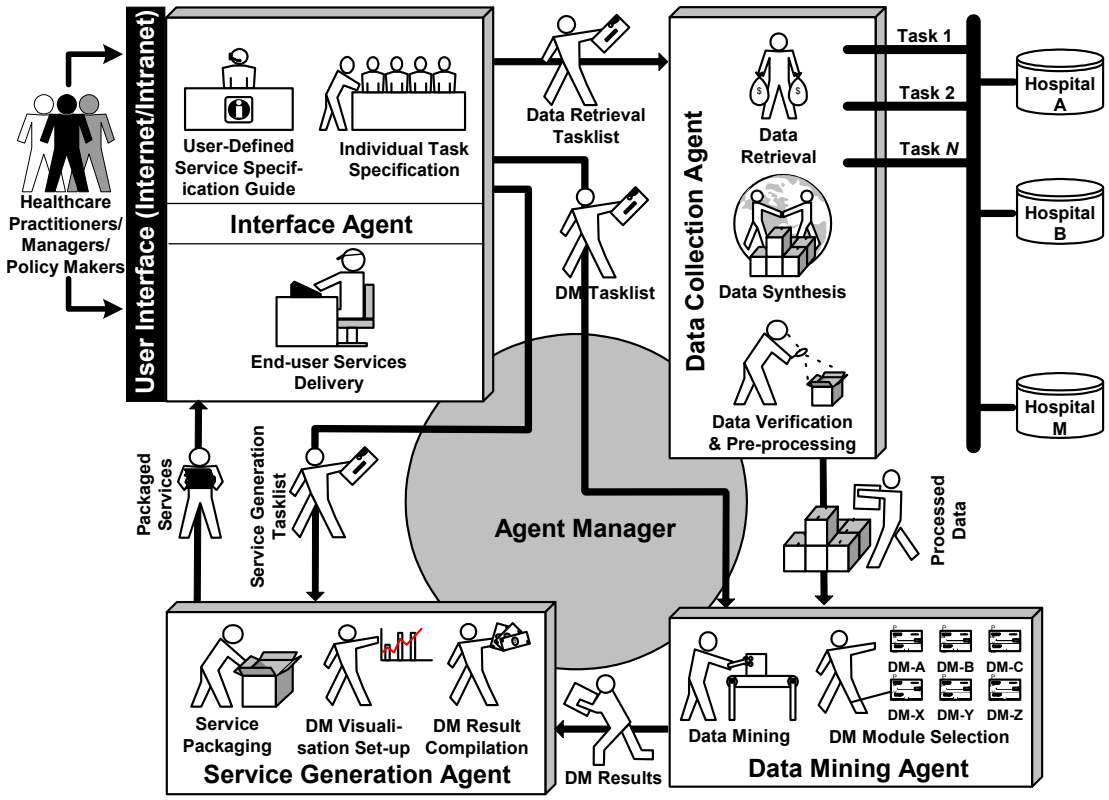

Figure 1. A sketch of our agent community—i.e. ADMI—which deals with the DM task given in scenario. Illutrated is the intrinsic behavior and inter-agent communication of the inherent intelligent agents. The shaded area represents the *Agent Manager*—a meta-agent responsible for coordinating all the agents.

3.2. 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 performs the following tasks: (a) establishing communication protocols for remote data access; (b) data selection and retrieval; and (c) heterogeneous data synthesis.

3.3. 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.

3.4. 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.

4. Conclusions

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* [5].

Currently, a prototype version of ADMI is under development and we are focusing on exploring ways of agent collaboration to realise a distributed DM info-structure.

5. References

[1] A.Babic, "Knowledge Discovery for Advanced Clinical data Management and Analysis", P.Kokol 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] 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.

[5] S.S.R.Abidi, "Applying Knowledge Discovery in Healthcare: An Info-Structure for Delivering Knowledge-Driven Strategic Services", P.Kokol et al. (Eds.) Medical Informatics Europe'99, Ljubljana, Amsterdam:IOS Press, pp. 453-456, 1999.