Optimizing Medical Data Quality Based on Multi-agent Web Service Framework

Ching-Seh Wu, Wei-Chun Chang, Sena Cebeci and Ibrahim Khoury
Department of Computer Science and Engineering, Oakland University, USA
{cwu, chang234, secebeci, iskhoury}@oakland.edu

Abstract—One of the important issues of the e-healthcare information systems is to precisely identify satisfied quality of the medical data from a distributed and heterogeneous environment. This paper proposes a multi-agent web service framework based on Service Oriented Architecture (SOA) for the optimization of medical data quality in the e-healthcare information system. Based on the design of the multi-agent web service framework, an evolutionary process for the dynamic optimization of the medical data quality is proposed and supported with experimental results by building up a case study. In dealing with complex and large-scale medical data requests, there is a foreseeable bottleneck of supporting technology in medical data selection. This framework can provide the optimized medical data quality and dramatically reduce medical errors caused by misjudgment of the e-healthcare Information System.

Index Terms—e-Healthcare, Medical Data Quality, Web Service

I. INTRODUCTION

It was reported that between 44,000 and 98,000 deaths occur annually as a consequence of medical errors within American hospitals alone [1] and the US National Association of Boards of Pharmacy reports that as many as 7,000 deaths occur in the US each year because of incorrect prescriptions [2]. Therefore, a great desire to improve access to new healthcare methods, and the challenge of delivering healthcare becomes significant nowadays. In an attempt to meet these great demands, health systems have increasingly looked at deploying information technology to scale resources, to reduce queues, to avoid errors and to provide modern treatments into remote communities.

Many medical information systems are proposed in the literature trying to assist in management and advising medical treatments to prevent from any type of medical errors. From the individualized care point of view, in order for clinicians to make the best diagnosis and decide on treatment all the relevant health information of the patient needs to be available and transparently accessible to them regardless of the location where it is stored. Moreover, computer-aided tools are now essential for interpreting patient-specific data in order to determine the most suitable therapy from the diagnosis, but existing systems lack collaborative ability because of employing different method to design [3].

Many researchers have been trying to apply service oriented architecture (SOA) to deal with the distributed environment for e-healthcare information system [6]. The objective of SOA is to provide better quality service to users. The new service is called web services [4]. Following the definitions and specifications of web service, any organization, company, or even individual developers who can deliver such functional entities can register and publish their service components to a Universal Description, Discovery, and Integration (UDDI) registry for public use. Web services can be as simple as a single transaction, e.g. the querying of a medical record, or more complex multi-services, e.g. supplying chain management systems from business to business (B2B), and many other [4]. However, current web service developments mostly focus on providing either a single service or at most a few. Focusing on single service without being prepared for complex and large-scale web services cause technological bottlenecks to develop. Therefore, in order to enhance service-oriented integration in distributed e-Healthcare environment, the collecting and composing of web service components for complex and large-scale web service applications need to be developed and improved.

In composing web services, both a single service component and a series of service components that can support large-scale tasks need to be found. Ko and Neches also point out that current web service research focuses only on developing mechanisms to describe and locate individual service components in a network environment [5].

In dynamic optimization of medical data quality, the information regarding suitable medical data service components need to be acquired from many medical data service providers whose components are registered in a UDDI registry repository. The next step is to negotiate with different medical data service providers in order to integrate suitable medical data components. The optimization of medical data selection is successful when multi-objectives set by a medical data service requester are met such as reliability of medical data component, results of diagnosis, and cycles of
consultation [8]. To evaluate web service composition, several aspects of the quality of service have been proposed, e.g. web service composition – Business Process Execution Language for Web Services (BPEL4WS) [7], web service coordination, web service transaction, web service security, and web service reliability.

This paper aims to apply the SOA of Web Service concepts specified above to put forward a model of multiple intelligent agents based assistance in improvement of medical data quality in the distributed e-Healthcare information system environment which is able to optimize the medical data web services according to data quality aspects. Furthermore, to improve accuracy of doctor’s diagnostic, many methods for Medical Diagnostic and Treatment Advice Systems have been developed to assist medical doctors in decision making such as rule based reasoning, fuzzy inference, neural network, and etc [8, 9, 10]. Intelligent Agent is another approach taken by researchers trying to assist in different domains such as business process, remote education service, and project management [11, 12, 13]. Our objectives of this research are to design and develop medical data quality models and to develop the methodologies and algorithms of our multi-agent framework to assist in monitoring and optimizing data quality for e-Healthcare information system.

In the following sections, we will first describe the preliminary aspects of our study focusing on medical data quality in terms of data extraction in section II. In section III, static and dynamic behavior of medical data quality models were designed and developed by using UML notations. These models will be implemented for healthcare intelligent agents to monitor and keep track of the medical data recording and extraction process in section IV. In section V, optimal medical data quality framework in distributed medical data environment is introduced and Evolutionary Computing is used when optimizing the data selection. Case study for the Breast Cancer disease is examined and indicated with experimental results using the Evolutionary Algorithm in section VI. Finally, Section VII concludes the paper.

II. PRELIMINARIES OF MEDICAL DATA QUALITY

Data quality refers to many different aspects. In Table 1, aspects of the data quality were grouped into two categories of dimensions, measurable dimension and intangible dimension. However, the main focus of the medical data quality in this research has been on the measurable Accuracy of data quality dimension. The accuracy of medical data in this study refers to the reality presentation of the medical data from data extraction process during the healthcare governance cycle specified in Figure 1. To receive an accurate set of medical data for healthcare consultation, this study has designed healthcare intelligent agents to monitor and track the data extraction process.

A. Healthcare Governance Cycle

In order to design intelligent agent to monitor and to keep track of medical data processing, the healthcare governance cycle is illustrated in Figure 1. Within the healthcare consultation, the General Practitioner (GP), such as a family doctor, uses a networked Healthcare Maintenance Organizations (HMO) to find relevant healthcare knowledge for the treatment.

<table>
<thead>
<tr>
<th>Table 1. Aspects of the Data Quality</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Measurable Dimensions for Data Quality</strong></td>
</tr>
<tr>
<td>Dimension</td>
</tr>
<tr>
<td>Accuracy</td>
</tr>
<tr>
<td>Integrity</td>
</tr>
<tr>
<td>Consistency</td>
</tr>
<tr>
<td>Completeness</td>
</tr>
<tr>
<td>Uniqueness</td>
</tr>
<tr>
<td>Accessibility</td>
</tr>
<tr>
<td>Precision</td>
</tr>
<tr>
<td>Timeliness</td>
</tr>
</tbody>
</table>

In the healthcare consultation, the General Practitioner (GP), such as a family doctor, uses a networked Healthcare Maintenance Organizations (HMO) to find relevant healthcare knowledge for the treatment.

Fig. 1. Healthcare Consultation Governance Cycle

The healthcare data from each consultation will be stored in medical database. The medical database records information in a concise format with compressed detail clinical coding regarding symptoms, diagnostic results, treatments, prescriptions, and other medical information for the consultation. When one of particular medical information is retrieved for further or next healthcare consultation/reference, the compressed medical data must be extracted to an understandable format for GPs. The feedback on the medical data quality will be conducted to improve the patient care for the next iteration of healthcare consultation.

The major concern of the medical data quality is drawn from the data extraction process. One of the major challenges in healthcare domain is the extraction of comprehensible knowledge from medical diagnosis data. Data accuracy and consistency must be maintained during the extraction process. In order to make sure that the data extraction process maintains a good quality of medical up-to-date information, medical data quality models are created for the further design of healthcare intelligent agents.
III. MODELING THE MEDICAL DATA QUALITY USING UML

A saying from Software Engineering said [14], “If you can model it, you can implement it.” We have designed the class diagram, the activity diagram, the use case diagram, and the sequence diagram for modeling the static view of the medical data quality and the dynamic behavior of the medical data extraction process using UML. By developing models, we are able to look into the details of the medical data recording/retrieval process as well as the data extraction process. This will help us design the multiple intelligent agents to monitor and track the data recording/retrieval and extraction processes to assist in medical data quality improvement.

Use case diagram is a tool for modeling the features and functions of an information system. The Use Case Diagram in Figure 2 shows that our system consists of medical data quality features such as data extraction, data migration, data cleaning, data integration, data processing and data analysis. Data analysis involves feedback and quality assessment. This Use Case Diagram is the first step toward the definition of the behavior of the medical data quality involved in a healthcare information process. For this particular study, we only focus on the data extraction of the medical data quality. The rest of medical data quality issues specified in Figure 2 has been reserved for future study and development.

The class diagram of medical data quality model in Figure 3 contains all classes/objects that associate with medical data processes and queries. Each class/object in the model was used to generate data quality metrics/attributes for intelligent agents to keep track and monitor. When data extraction process is conducted, the medical information in classes/objects will be collected in medical knowledge base for inference conducted by intelligent agents.

![Fig.3. Medical Quality Modeling – Class Diagram](image)

The activity diagram in Figure 4 models the activities and tasks involved in the medical data extraction process. These key activities including using hospital query language (HQL) for hospital information system, medical data recording process, and looping process for data update. The activity diagram helps to design the internal monitoring process of healthcare intelligent agents.

The sequence diagram of medical data model in Figure 5 shows the process sequence of medical data extraction. The sequence diagram enhances the process definition from the activity diagram. Healthcare intelligent agent uses these two models to identify quality items to be monitored.

![Fig.2. Medical Data Quality Modeling – Use Case Diagram](image)
IV. OPEN DESIGN OF INTELLIGENT AGENT PROTOTYPE

Once both dynamic and static models for the medical data quality have been designed, the intelligent agent was developed according to the medical quality models to monitor and to keep track of medical data recording and processing. Our design of intelligent agent is an open and collaborative infrastructure so that each agent has the same structure enabling communications with each other. An agent, as illustrated in Figure 6, consists of three service components, collaboration service, quality monitor service, and reporting service as described in Figure 8. The collaboration service enables agents to plug & play medical web forms for portable medical records and to plug & play data workflows, medical protocols, and clinical guidelines in a distributed heterogeneous medical information environment, and enables information exchange service among intelligent agents. There are two existing methods that can be used to implement intelligent agents, Java Expert System Shell (JESS) and Java Agent Development Framework (JADE). The interior implementation of services for an agent was carried out by using JESS. The exterior communication behavior in a distributed e-healthcare environment was carried out by using JADE. In general, healthcare intelligent agents have been developed to assist e-healthcare information system in medical data quality improvement activities:

- update medical knowledge base
- define criteria for healthcare data query
- determine if a threshold value of data quality has been reached
- optimize data quality for the accurate diagnosis
- keep track of patient’s healthcare profile
- communicate with other agents

JESS was used to develop and implement the interior inference process of intelligent agents as specified in Figure 7. The intelligent agent can take inputs from healthcare experts and transfer the inputs into healthcare knowledge for inference engine to make consultation judgments. Healthcare intelligent agent was developed with the feature described in Figure 8. It is able to take inputs from medical events such as patient’s medical history, diagnosis knowledge from experts, and medical symptoms. The interior features include update medical knowledge, define criteria for hospital queries, determine if the data quality threshold value has been reached, keep track of patient’s medical profiles, and communicate with other agents.
Fig. 7. Intelligent Agent Inference Structure

Fig. 8. A Healthcare Intelligent Agent

V. OPTIMIZATION OF MEDICAL DATA QUALITY

One of the most important concerns of this study is the medical data selection for quality improvement over a distributed e-healthcare information environment. The foundation of satisfying data quality over the distributed medical data environment compiles the analysis and construction of medical data service workflow, the automation of composing/optimizing suitable medical data Web Service components, and medical data Web Service component reusability. To satisfy data quality criteria, we proposed a framework (see Figure 9) where we integrated intelligent agent, a medical data repository section and several modules into the Service Oriented Architecture (SOA).

Evolutionary Algorithms (EAs) have been applied as the searching algorithms to search the optimal medical data in the distributed e-healthcare information environment as specified in Figure 9. “Survival of the fittest” [15] is a principle in the natural environment which is used in the medical data selection algorithm to generate survivors, the optimal data selection in the distributed healthcare environment.

The original principles of the EC theory are based on Darwin’s theory of natural selection to solve real world problems [6]. EAs have been successfully applied in optimizing the solutions for a variety of domains [6]. The strength of EC techniques comes from the stochastic strategy of search operators.

The major components in EC are search operators acting on a population of chromosomes. EC was developed to solve complex problems, which were not easy to solve by existing algorithms [6, 7]. The method utilized in the algorithm to progress the search from ancestors to offspring is the collective learning process; species information is collected during the evolutionary process, and the offspring that inherit good genes from parents survive the competition. This is the first characteristic of EAs. Next, the generation of descendants is handled by the search operators, crossover and mutation; which explore variations in species information in order to generate offspring. Crossover operators exchange information between mating partners. On the other hand, a mutation operator, which mutates a single gene with very small probability, is used to change the genetic material in an individual. Finally, the third characteristic that defines EAs is the evaluation scheme, which is used to decide who the survivor is. The evaluation scheme is the most diverse characteristic of the three due to the different objectives used to select the different solutions needed in different domains. The evaluation scheme can be as simple as good or bad, a binary decision; or as complex as nonlinear using multiple mathematical equations to assess trade-offs between multiple objectives.

For this study EC techniques provided stochastic searching techniques aimed at global optimization. Global optimization searches for the optimal performance of solutions in the objective space. A general global optimization problem can be defined as follows:

$$f_*(x) = \min_{x \in D} f(x)$$

subject to $c(x)$
Where $f_d(x)$ is the global optimization in objective space when determining the minimum of the function $f(x)$; $x$ is a vector of variables which lies in the feasible region $\Omega$, any $x$ in $\Omega$ defines a feasible solution in which $x$ conforms to the constraints $c(x)$. A similar definition can also be applied to the maximization of objective functions.

1) Collecting the medical data of component registrants: the size of medical data searching space is decided by the number of component registrants collected from available UDDI registries. Therefore, it is very important to obtain the information of all available medical data locations/components from component registration agents. The information regarding the description of service components can be collected from a component library as specified in [16]. The communication protocol is based on a set of API message (i.e., UDDI 3.0 and up).

2) Modeling medical data resources from different providers: medical data service components are classified and constructed into database tables based on the functionalities and characteristics of medical data service requested. The work flow of the medical data service can be modeled by using a scenario-based method that is used in previous sections to describe the task steps required to accomplish the completion of medical data web service applications[15].

3) Applying the sequence of medical data web service composition and chromosome encoding/decoding: the task sequence of medical data web services that are needed to be optimized is defined. A sub-task service in a task sequence can be defined as:

$$\{ \text{component}_{ji}, \text{sub-task}_j \}$$

where it is assumed that one sub-task can be completed by a medical data service component. By utilizing the collected information of medical data component registrants, a web service task sequence is transformed into a binary string, i.e. encoding a quality solution into a chromosome. The chromosome mapping mechanism utilizes a hierarchical structure [15] for an encoding/decoding task sequence and chromosome.

4) Quality Fitness Assessment: To evaluate the quality of medical data optimization, multi-parameters or attributes are used in the metrics to evaluate performance and quality.

The metric measurement focuses on different aspects that data quality criteria require. Such measurement is a key element of evaluating the performance and quality of medical data optimization.

### VI. Case Study

In this section, we propose a case study for e-healthcare subsequently used throughout the paper. To demonstrate Evolution Algorithm, Breast Cancer Web Service Task Sequence is selected in this case study. The main goal of this case study is to find the optimal solutions for diagnostics, treatments and alternative treatments according to multi-objective medical data quality metrics. To indicate the efficiency of EA for optimizing data quality we implemented the algorithm using MATLAB platform.

Figure 11 shows the web service task sequence of the Breast Cancer. Following a physical examination, wherein the patient has been found to have Breast Cancer, the next step in the sequence, will be choosing a series of tests that the patient will undergo to provide treatment [19].

Table 2 shows the test, treatment and alternative treatment types in Breast Cancer. Based on test results, treatments and alternative treatments are decided upon by the doctor.

Fig.11. Breast Cancer Web Service Task Sequence
We focus on a set of data quality dimensions which are provided by intelligent agents, namely: accuracy, consistency, completeness, and timeliness, which constitute the focus of the majority of authors [17, 18].

Table 2 Tests and Treatment Types

<table>
<thead>
<tr>
<th>Tests</th>
<th>Treatments</th>
<th>Alternative Treatments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mammogram</td>
<td>Radiation Therapy</td>
<td>Bioflavonoid</td>
</tr>
<tr>
<td>MRI</td>
<td>Surgery</td>
<td>Mineral Supplement</td>
</tr>
<tr>
<td>X-Ray</td>
<td>Hormonal Therapy</td>
<td>Vitamin Supplement</td>
</tr>
<tr>
<td>Ultrasound</td>
<td>Chemotherapy</td>
<td>Herbal Supplement</td>
</tr>
<tr>
<td></td>
<td>Targeted Therapy</td>
<td>Exercise</td>
</tr>
</tbody>
</table>

Table 2 Tests and Treatment Types

In order to evaluate performance and quality of medical data optimization, data quality dimensions used in metrics. Definitions for quality requirements that are used in the experimental results section are as follows:

- **Accuracy**: the data should be presented as reality or verifiable medical resources.
- **Completeness**: all specific information has to be represented as complete.
- **Consistency**: data should be represented without repetition.
- **Timeliness**: medical data should be stored and updated constantly.

Our target is providing the optimal solution by using these four quality requirements: the assumption is having the accuracy, completeness, and consistencies are at a maximum where as timeliness is at a minimum.

Then, we return the optimization of medical data quality section of the paper to illustrate how our proposed approach should work in this case study. In the optimized data quality framework in Figure 9, it is shown that the intelligent agent reports the data quality metrics to the EA and EA finds the optimal solution for Breast Cancer Task Sequence. Intelligent agent duty in the framework is to keep track and verify the data extraction is completed regarding to the optimal task sequence provided by the EA.

A. **Experimental Results**

This section presents several experimental results to indicate the efficiency of the multi-agent framework based on web service in medical data quality integrated with the Evolutionary Algorithm.

Figure 12 shows the efficiency of the algorithm in maximizing the 3-objectives solution space: accuracy, consistency, and completeness. Each points represent the combination of service components to complete the web service task sequence. Our algorithm reaches the fittest task sequence after 12 generations which is indicated with an arrow.

Figure 13 and Figure 14 show how Evolutionary Algorithm can solve multi-objective problems by using combination of different quality metrics with different target points. Such as: Maximizing accuracy, consistency and completeness, where as minimizing timeliness.

Figure 15 shows the efficiency of the EA to optimize the solution is converged within 12 generations. This convergence indicates the applicability of the EC algorithms in optimizing web service task sequence. The algorithm finds the optimized solution and also gives alternative solutions for test, treatment and alternative treatment. The simulation was tested through set of data quality metrics randomly generated data set by MATLAB, the results demonstrated that the metric
The design of the intelligent agent enables the intelligent agent to handle the problem of data selection and quality optimization in a distributed e-Healthcare environment, evolution computing algorithm was integrated into the Service Oriented Architecture of Web Service. In SOA, the healthcare intelligent agent also plays a major role as the service agent for medical data registration service and data requesting service. This multi-agent framework has been developed using Java Expert System Shell (JESS) and Java Agent Development Framework (JADE). The system will be practically deployed and integrated with e-Healthcare information systems for our local hospitals.

**REFERENCES**


