Rule Based Classification to Detect Malnutrition in Children

Xu Dezhi
School of Information Science & Engineering
Central South University
Changsha, China
hunan.xu@mail.csu.edu.cn

Gamage Upeksha Ganegoda
School of Information Science & Engineering
Central South University
Changsha, China
upeksha.ganegoda@gmail.com

***Abstract* – Data mining is an area which used in vast field of areas. Rule based classification is one of the sub areas in data mining. From this paper it will describe how rule based classification is used alone with Agent Technology to detect malnutrition in children. This proposed system is implemented as an e-government system. Further it will try to research whether there is connection between number of rules which is used with the optimality of the final decision.**

Keywords- data mining, rule based classification, e-government, malnutrition, multi agent system

1.0 INTRODUCTION

Data mining, a branch of computer science and artificial intelligence, is the process of extracting patterns from data. Data mining is seen as an increasingly important tool by modern business to transform data into business intelligence giving an informational advantage. It is currently used in a wide range of areas, such as marketing, health, surveillance, fraud detection, and scientific discovery. This has been improved by other discoveries in computer science, such as neural networks, clustering, genetic algorithms, decision trees and support vector machines. Hence data mining apply these methods to data with the intention of uncovering hidden patterns. The technology of data mining is widely used in various fields. E-Government is a grand new domain in recent years. When E-Government system use to process data, it need to choose what data is useful and what kind of new information which can get from the log file or from the database [18]. As e-government is the use of Information Technology to free movement of information to overcome the physical bounds of traditional paper and physical based systems [8, 3]. Hence data mining techniques can be used to improve e-government systems. The current situation of e-government in Sri Lanka is still in the developing stage and there are minimum interactions between health sectors. During the recent past, Sri Lanka has improved in nutrition, underweight rates declined by almost 25%, long term under nutrition declined by about 75% [11]. Hence this research is carried out to review the current status of e-government in Sri Lanka and to provide a way to reduce malnutrition in children by introducing a new system which will function under the e-government initiatives in Sri Lanka [12] by using Agent Technology and Data mining technique such as rule based classification. During the research it has test whether there is an effect on number of rules associated to make the final answer optimal. e-SriLanka is the main e-government project in Sri Lanka. This project was embarked in 2002 and was handled by the Information and Communication Technology Agency (ICTA) in Sri Lanka [7]. 'Nanasala' is one of the projects carried out in the e-society programme. Currently through Nanasala centers it is promoting e-government projects among citizens as well as health consultation. For an example Tele-Genetic project provide health consultation for genetic to rural areas through the Nanasala centers [7]. Interoperability is also an important factor in e-government sector [16]. According to the UNDP e-government Interoperability Overview [4] the technical perspective of interoperability is "the ability of two or more government Information and Communication Technology (ICT) systems or components to meaningfully and seamlessly exchange information and use the information that has been exchanged." As the description of Government Interoperability Framework (GIF) in UNDP e-government Interoperability Review [15] GIF can be developed according to the layered model as well as from the use of life cycle method. Sri Lanka has also created their own interoperability framework called Lanka Interoperability Framework (LIFe) [10, 1]. Currently various projects implemented in Sri Lanka under the initiative of e-government. The project called the medical information portal will provide static text based and multimedia based (video streaming) health education content in Sinhala and Tamil to citizens. From the remote medical consultation system (Vidusuwa) it handles the challengers faced by the patients in Sri Lanka with regard to inequality of resource distribution within the existing eHealth infrastructure [6].

The section 2 describes an overview of rule based classification and how it is used in health sector. The section 3 will describe the enhanced system of Multi Agent System to Reduce Malnutrition with the use of rule base classification technique for detection purpose. It will describe the features of the system and whether there is any connection between number of rules and the final answer.

Project 60970096 supported by National Natural Science Foundation of China

2.0 OVERVIEW OF RULE BASED CLASSIFICATION

Classification is an important field of data mining problems. Given a set of labeled training examples the classification task constructs a classifier. A classifier is a global model which is used to predict the class label for data objects that are unlabeled. Many approaches have been proposed for the classification problem. Among them, rule-induction, as- associative and instance-centric approaches have been closely integrated with constraint-based data mining. A rule based classifier is a technique for classifying records using a collection of "if… then…" rules. The rules for the model are represented in a disjunctive normal form, $R = (r_1 \ V \ r_2 \ V \ \ldots. \ R_k)$, where R is known as the rule set and r_i's are the classification rules or disjuncts [14].

Each classification rule is expressed in the following way:

$$r\ i:\ (\text{condition } i) \longrightarrow y\ i$$

The left- hand side of the rule is called the rule antecedent or precondition. It contains a conjunction of attribute tests:

$$\text{Condition } i = (A_1 \text{ op } v_1) \wedge (A_2 \text{ op } v_2) \wedge \ldots. (A_k \text{ op } v_k)$$

Where (Aj, vj) is an attribute-value pair and op is a comparison operator chosen from the set $\{=, \neq, \leq, \geq, >, <\}$. Each attribute test (Aj op vj) is known as a conjunct. The right-hand side of the rule is called the rule consequent, which contains the predicted class yi. A rule r covers a record x if the precondition of r matches the attributes of x. r is also said to be fired or triggered whenever it covers a given record.

The quality of a classification rule will be evaluated using measures such as coverage and accuracy. Given a data set D and a classification rule r: A→y, the coverage of the rule is defined as the fraction of records in D that trigger the rule r. accuracy or confidence factor is defined as the fraction of records triggered by r whose class labels are equal to y. The formal definitions of these measures are

$$\text{Coverage}(r) = \frac{|A|}{|D|}$$

$$\text{Accuracy}(r) = \frac{|A \cap y|}{|A|}$$

Where |A| is the number of records that satisfy the rule antecedent, $|A \cap y|$ is the number of records that satisfy both the antecedent and consequent, and |D| is the total number of records.

In rule base classification technique rule extraction is one of the major sections. There are two broad classes of methods for extracting classification rules: 1) direct methods, which extract classification rules directly from data, and 2) indirect methods, which extract classification rules from other classification models, such as decision trees and neural networks.

In direct method there are several methods which can use to extract the rules from the data set. From the sequential covering algorithm rules are generated in a greedy fashion based on a certain evaluation measure. The algorithm extracts the rules one class at a time from the data set. The criterion for deciding which class should be generated first depends on factors such as class prevalence or cost of misclassifying records from a given class.

Learn one rule function will extract a classification rule that covers many of the positive examples none of the negative examples in the training set. It addresses the exponential search problem by growing the rules in a greedy fashion. It generates an initial rule r and keeps refining the rule until a certain stopping criterion is met. The rule is then pruned to improve its generalization error [14].

Final method is rule growing strategy. In this method there are two common strategies for growing a classification rule: general-to-specific or specific-to-general. Under the general-to-specific strategy, an initial rule r: {} → y is created, where the left-hand side is an empty set and the right-hand side contains the target class. The rule has poor quality because it covers all the examples in the training set. For the specific-to-general strategy, one of the positive examples is randomly chosen as the initial seed or the rule-growing process. During the refinement step, the rule is generalized by removing one of its conjuncts so that it can cover more positive examples. By considering these factors in the rule-based classification technique will consider the accuracy, coverage, error rate, sensitivity, specificity of the rules.

Some research carried out to check the relationship between Health-Related Quality of Life (HRQOL) and the International Classification of Functioning Disability and Health (ICF) represent two different perspectives from which to look at functioning and health [2].

By considering these factors in the rule-based classification technique it will consider the accuracy, coverage, error rate, sensitivity, specificity of the rules. But from this research it will consider whether there is a connection between number of rules which used with the accuracy of the final decision. Therefore during the research it will consider how the final decision will change according to the number of rules which use to make the final decision.

3.0 Introduce Agent Technology to overcome malnutrition issues in the e-government background

According to current research in the world, there is a mobile multi-agent based system to monitor e-health [5]. As well as agents developed to control unknown diseases [13] and Alzheimer patients [9]. By analyzing the current situation of e-government and the nutritional status of Sri Lanka, it is clearly highlighted that there are no ongoing projects which are initiated by e-government to reduce malnutrition in Sri Lanka. In addition, currently e-government in Sri Lanka does not use Agent Technology for health projects. Due to these reasons from this research propose to use Agent Technology to overcome this situation [17].

A. Multi-Agent System to Reduce Malnutrition Framework

Multi-Agent System to Reduce Malnutrition (MASRM) is developed under the influence of Java Agent Development Framework (JADE) runtime environment. Therefore each instance of the runtime environment, the "Container" has corresponding Agent Management System (AMS), Directory Facilitator (DF). AMS is responsible to provide unique names to each agent in the system and creation and termination of the agent. DF is the agent who provides a yellow pages service by means of which an agent can find other agents to achieve a specific service. Remote Monitoring Agent (RMA) is a system tool which represents all the actions in the platform in a Graphical User Interface (GUI). Therefore it consist options to control an agent's life cycle, of an agent which is done by the AMS. The Message Transport System is also called the Agent Intercommunication Channel (AIC), is the software component controlling the exchange of all messages within the platform, including messages to and from remote platforms. It will make sure that all the messages have the Agent Communication Language (ACL) format which is issued by FIPA.

B. System Architecture

MASRM was developed on top of JADE, which is suitable to operate in a heterogeneous, networked environment such as the Internet to provide wide-area health detection and health information service to the citizen in the country. The framework further improved by introducing another agent called Resource Agent to the framework. Fig. 1the system consists of six types of architectural components which will help to identify malnutrition children in the real world: (1) User Identify Agent (2) Detection Agent (3) Advice Agent (4) Knowledge – based data server (5) Medical Agent (6) Resource Agent. It illustrates the structure of a container in the system. Another agent called Reporting Agent will only function in the main container of the system.

1) User Identify Agent

This will act as a stationary agent which will identify the users. When the users input login information through the web interface it will activate the user identify agent (UIA). Hence it will receive the request from the user interface. It will authenticate the user before starting the agent service and will arrange the data which is passed by the detection agent, advice agent or from the medical agent before passing back to the user interface.

2) Detection Agent

Detection Agent (DA) will be activated only when the user want to detect the nutrition status of their child. In this case the user will be the father or mother of the child. Hence it will collect the data given by the UIA for the particular user. Then this data will be analyzed with the existing knowledge which is already in the knowledge base and detect whether the child has malnutrition or not. Therefore DA will act as a data analysis engine in the system. If the agent detects malnutrition, then it will notify the corresponding medical agent as well as to the UIA.

3) Advice Agent

Advice Agent (AA) will be activated when the user want to get nutrition advice from the system. If the user selects the option then he/she have to answer a set of questions which is prompt by the AA. By analyzing the answers given by the users, agent will provide nutrition advice. Hence this will be another data analysis engine in the system.

4) Knowledge-based Data Server

It will store the knowledge about the users, parameters to check malnutrition as well as different nutrition advice for different situations. It will interact with the detection agent, advice agent as well as with the medical agent. Knowledge-based data server can be updated with new knowledge also.

5) Medical Agent

If the system detects malnutrition or an unusual nutrition status then, the details about the particular child will be forwarded to the medical agent (MA). Then the MA will notify the situation to a real physician for further action. As well as MA will notify the status of malnutrition children who are registered under the physician periodically.

6) Resource Agent

Resource agent (RA) is a stationary agent which access the local data reposity. It is responsible of accessing the knowledge base and retrieves or store necessary information. Therefore resource agent acts as an interface to bridge the users to the knowledge base.

7) Data Service Environment

All the knowledge and data related to the system is handling within the data service environment. It consists of a repository, which will store and maintain the information about the agents' status, user information, causes and various advices for malnutrition.

8) Agent Service Environment

Within this environment all the agents in the system will function. It provides base services which the agent-based applications may utilize such as inter-agent communication. This will be done by using a dedicated Agent API.

9) Reporting Agent

This will be a mobile agent which is situated only in the main container. The main responsibility is to notify each midwife whether the responsible children are in malnutrition or not. Hence to collect all the information reporting agent will move from container to container and reporting back to the main container.

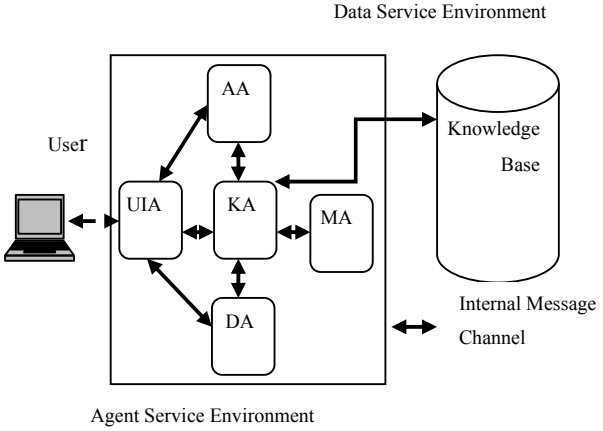

Figure 1. Enhanced Multi-Agent System to Reduce Malnutrition Architecture

4.0 USER SCENARIOS

This section will describe three main user scenarios in the system.

A. *Scenario 1: detection of malnutrition*

The process will start when a user i.e. a parent of a child logs into the system and request the system to check nutrition status of their child. This will activate the UIA. Then the user input child information into the system. As shown in the Fig. 2 the user needs to fill specific information about the child. Then UIA will give the information to the DA and KA. KA will send a copy of the information to the knowledge base. At the same time KA will forward the detection parameters to DA according to the DA request. Finally DA will detect whether the child consist malnutrition or not. Detection will be carried by using the rule based classification techniques which will describe in the next section. If malnutrition is detected then the status of the child is informed to the corresponding physician through the MA. MA will send the message as an email to the physician. Lastly the decision of the DA is forward to KA and it will update the knowledge base and will inform the user through the UIA.

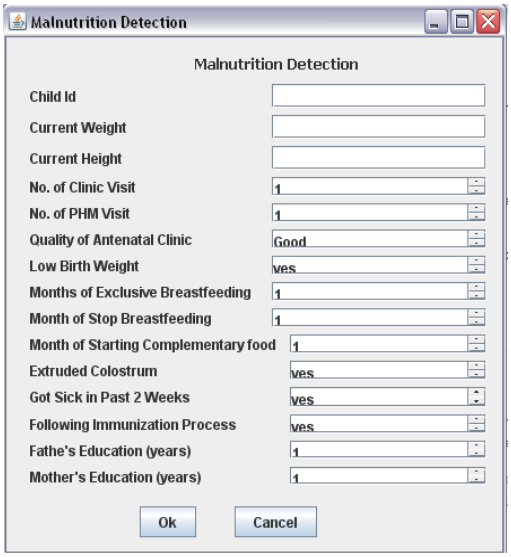

Figure 2. Screen of Malnutrition Detection

B. *Scenario 2: advice on nut*

A user (i.e. user will be the parent) requests nutrition advice from the system after logging into the system. This request will activate the AA and AA will forward set of questions as shown in Fig. 3 which should be answered by the parent. A copy of the user input will be saved in the knowledge base with the help of KA. Then AA will analyze the user information and request correct nutrition information from the KA. Hence KA will fetch the information from the knowledge base and forwarded to AA. After final analysis nutrition advice of the situation will be forwarded to UIA through KA. UIA will present the data to the user.

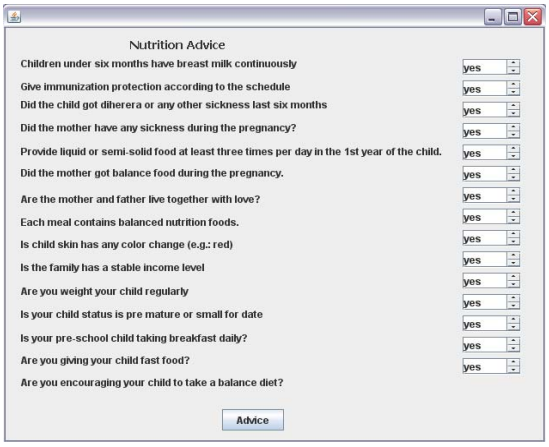

Figure 3. Screen of Nutrition Advice

C. Scenario 3: reporting malnutrition details

This scenario functions only in the main container which is situated in head office by the Reporting Agent. It will be monthly activated automatically. At the beginning of each month this mobile agent will move from one container to another and calculate the number of malnutrition children in each district. As one container represent one district. After collecting all the information the reporting agent will generate a monthly report on malnutrition children in Sri Lanka during that period according to specific midwife. Hence midwives also can log in to the system to check whether the children under her responsibility are malnourished or not.

5.0 IMPLEMENTATION AND EXPERIMENT

The proposed enhanced framework will be implemented as a web- based system. Therefore the users are able to login to the system via the Internet. Parents and midwives will be the main users of the system. Midwives will be the direct physicians who are involved in child growth. In Sri Lanka each district has a main Medical Officer of Health (MOH). Many midwives are working under each MOH. Each midwife take care the nutrition status of several children under five years old. Therefore the system will be implemented in such way that each MOH is represented by one container. Each container consist five agents. Namely: User Identify Agent, Detection Agent, Advice Agent, Medical Agent and Resource Agent. The main container will be running in the Head Office. Hence any container should register with the main container during the startup. The container which run in the head office has a special mobile agent called Reporting Agent. Reporting the number of malnutrition children in each district to the head office will be the main activity of the reporting agent. Each container maintains a local knowledge-base data server. Knowledge-base data server will be implemented using Oracle 11g. All the data servers will maintain same internal data structure. But the information about each MOH will be stored in each data server. Citizens can access this system through the Nanasala centers in the district.

Detection Agent will be enhanced by introducing rule based classification techniques. To implement the experiment the database consists of 200 records. From that 150 records will be used as training set and 50 records as test set. According to the questions given to the users a set of rules will be generated. Then for each rule, (1) likelihood ratio value will be calculated. Then the system will select the best 35 rules according to the likelihood value. By using the 35 rules it will decide the final answers of the test set. Therefore it can compare the actual value with the obtain value. Above mention steps will be carried out for best 40, 45 and 50 rules as well. Finally a graph will be generated by showing Y axis number of correct records and X axis number of rules.

$$R = 2\sum_{i=1}^{k} f_i \log\left(\frac{f_i}{e_i}\right) \quad (1)$$

6.0 RESULT DISCUSSION

The final decision will be change according to the number of rules which have used to make the final decision. Due to this reason the final decision of the research is that there is a connection between number of rules and the optimality of the final decision.

7.0 CONCLUSION

Through this research it is highlighted that even Sri Lanka e-government initiative has been expanded to some extent, there are a lack of health related projects. Hence I have further enhanced the MASRM system which will be developed under the e-government initiative to provide advice regarding nutrition as well as to provide an easy way for citizens to check nutrition status of their child. From this research paper it has further improve the MASRM system by in cooperating rule based classification technique to detect malnutrition. Further from this research it is highlighted that there is an effect on number of rules which is used to make the final decision with the optimality of the final decision.

ACKNOWLEDGMENT

This research is under Project 60970096 supported by National Natural Science Foundation of China. As well as I wish to thank all the resource personal in World Health Organization, Ministry of Health, ICTA and University of Moratuwa in Sri Lanka by providing necessary information for my research.

REFERENCE

[1] Agency, I. a. (2008). "Addendum to the Lanka Interoperability Framework (LIFe)." Colombo: Information and Communication Technology Agency .

[2] Alarcos Cieza, G. S. (2005). "Content comparison of health-related quality of life (HRQOL) instruments based on the international classification of functioning, disability and health (ICF)". Quality of Life Research (2005) 14 (pp. 1225–1237). Springer 2005.

[3] Babita, G., Subhasish, D., & Atual, G. (2008). "Adoption of ICT in a government organization in a development country: An empirical study." Stratefic Information Systems , 140-154. in press

[4] Bangkok, U. R. (2007). “e-Government Interoperability Overview.” Thailand: United Nations Development Programme.
[5] Chuan, J. S. (2008). “Mobile multi-agent based, distributed information platform (MADIP) for wide-area e-health monitoring.” Computers in Industry , 55-68. in press
[6] Dr. Keith R. P. Chapman MD, D. S. (2010). “ViduSuwa – Electronic Distant Healing: A Patient Centric Telemedicine Solution in Sri Lanka.” Sri Lanka Journal of Bio-Medical Informatics 2010 , 63-75. in press
[7] Dr. V.H.W Dissanayake, D. N. (2010). “Tele-Genetics: Using Low Cost Internet Technology to Provide Genetic Consultations to Rural Areas in Sri Lanka.” Sri Lanka Journal of Bio-Medical Informatics , 1(1):46-48.
[8] Gupta, B., Dasgupta, S., & Gupta, A. (2008). “Adoption of ICT in a government organization in a developing country: An empirical study.” Stategic Information SYstems . in press
[9] Juan M., C., Javier, B., Yanira de, P., & Dante I., T. (2008). “Intelligent envirnment for monitoring Alzheimer patients, agent technology for health care.” Decision Support Systems , 382-396. in press
[10] Lanka, I. a. (6th January 2006). “Lanka Interoperability Framework.” Colombo: Ministry of Public Administration and Home Affairs.
[11] Malnutrition in Sri Lanka. (n.d.). Retrieved May 25, 2009, from Unicef: http://www.unicef.org/srilanka/activities_1667.htm
[12] Meera Shekar, A. S. (2007). “Malnutrition in Sri Lanka: Scale, Scope, Causes, and Potential Response.” Human Development Unit, Document of the World Bank .
[13] Mihaela, U., Maja, H., & Elizabeth, C. (2006). “Soft computing agents for e-Health in application to the research and control of unknown diseases.” Information Sciences , 1190-1214. in press
[14] Pang Ning Tan, M. S. (2006). “Introduction to Data Mining” China P. R: Pearson Education Asia Ltd.
[15] UNDP. (2007). “A Review of Government Interoperability Frameworks in Selected Countries.” Bangkok: United Nations Development Programme.
[16] Vassilios, P., Konstantinos, T., & Sotirios K., G. (2009). “Model-driven eGovernment interoperability:” A review of the state of the art. Computer Standards & Interfaces , 613-628. in press
[17] Xu Dezhi, G. U. (2010). “Multi Agent System to Reduce Malnutrition (MASRM) in Children.” International Conference on Computational Intelligence and Software Engineering (pp. 70-73). Wuhan, China: ISBN 978-1-4244-5392-4, IEEE Catalog Number: CFP1026H-CDR.
[18] Yilei Wang, H. P. (2007). “The Data Mining of the E-Government on the Basis on Fuzzy Logic”. International Conference on Integration Technology (pp. 774 - 777). Shenzhen, China: Proceedings of the 2007 IEEE.