10

Data Mining in Ubiquitous Healthcare

Viswanathan, Whangbo and Yang
Carnegie Mellon University, Adelaide,
Australia

1. Introduction

Ubiquitous healthcare is the next step in the integration of information technology with healthcare services and refers to the access to healthcare services at any time and any place for individual consumers through mobile computing technology. Further, ubiquitous healthcare is able to provide enhanced services for patient management such as services that collect patients' data real-time and provide health information by analyzing the data using biomedical signal measurement instruments, which can be carried anytime, anywhere and by everyone online as well as offline.

The emergence of these tremendous data sets creates a growing need for analyzing them across geographical lines using distributed and parallel systems. Implementations of data mining techniques on high-performance distributed computing platforms are moving away from centralized computing models for both technical and organizational reasons (Kumar & Kantardzic, 2006).

In this paper, we present and discuss the designed prototype for a ubiquitous healthcare system that will provide advanced patient monitoring and health services. Subsequently we introduce and present empirical analysis of a preliminary distributed data mining system. The integration of such a distributed mining system is studied in the context of the decision support framework for our ubiquitous healthcare system.

2. Ubiquitous healthcare initiatives

A growing number of ubiquitous healthcare projects are being pursued by large enterprises owning healthcare related companies and government bodies. MobiHealth project (MobiHealth, 2004) is a mobile healthcare project supported by the EC with countries such as Netherlands, Germany, Spain and Sweden participating in it, and companies such as Philips and HP are providing technical support. EliteCare, is an elderly care system developed in the USA that monitors patients using various sensors and provides emergency and health information services. Tele-monitoring service is being developed by the Philips Medical system, where centers analyze data that is collected from homes and transmitted by biomedical signal collection devices, and provide health management and related information. CodeBlue is a sensor network based healthcare system being developed to treat and deal with emergencies, rehabilitation of stroke patients, and in general, to use health signal data in addition to hospital records in real time treatment decisions. The UbiMon (Kristof Van Laerhoven et al., 2004) project which stands for Ubiquitous Monitoring Environment for Wearable and Implantable Sensors is studying mobile monitoring using sensors and real-time biomedical data collection for long time trend analyses. The Smart Medical Home project developed at the University of Rochester in New York aims to develop a fully integrated personal health system with ubiquitous technology based on infrared and bio sensors, computers, video cameras and other devices. Sensor data is collected and transmitted to a center for further analysis and preventive care.

There are several ubiquitous challenges in the development of such healthcare frameworks and systems. These include:

- issues of security and privacy related to information transfer through unsecured infrastructure, potentially lost or stolen devices, legal enforcement and other scenarios;
- determining current context and user activity in real-time and locating context dependent information such as automatic discovery of services based on user health needs;
- development of low-power sensors to monitor user context and health condition;
- information management through development of techniques to collect, filter, analyze and store the potentially vast quantities of data from widespread patient monitoring and applying privacy preserving data mining at several levels;
- simple patient interaction systems to provide guidance, feedback and access to medical advice in acute situations;
- Adaptable network infrastructures to support large-scale monitoring, as well as real-time response from medical personnel or intelligent agents.;
- integration of specialized local u-Health architectures for unified data access and connection to National grids;

3. U-healthcare system framework

The components of the ubiquitous system prototype are summarized in this section. A system user in this paper refers to a patient who has a contract with a provider to use the ubiquitous healthcare services and regularly receives medical treatment at a hospital. Fig. 1 shows an overview of the ubiquitous healthcare service framework as suggested in this paper.

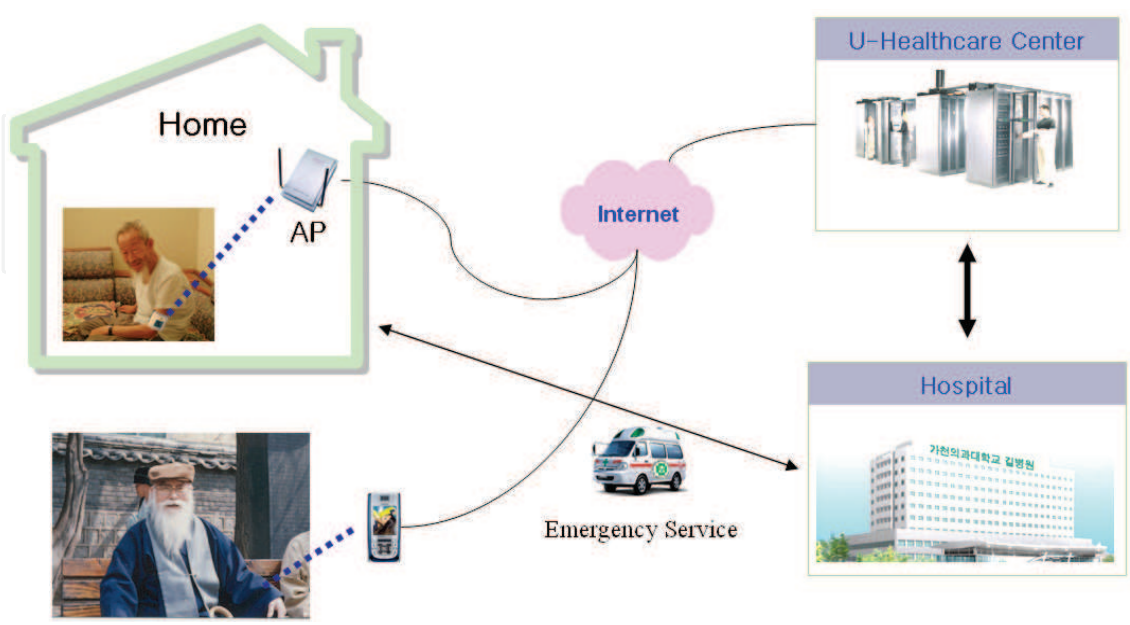


Fig. 1. Ubiquitous Healthcare Framework

The user wears a sensory device, provided by the hospital, on his wrist. The sensor regularly transmits collected data to a healthcare center through networking or mobile devices, and the transmitted data is stored at the u-healthcare center. In the center, monitoring staff are stationed to answer the user' queries, monitor his biomedical signals, and call an emergency service or visit the patient to check his status when an abnormal pattern is detected. The hospital monitors the collected data and judges the patient's status using the collected biomedical signals in his periodic check-up.

3.1 Biomedical signal collection and transmission

A wrist sensor is used to collect biomedical signals. The wrist sensor, attached to a user's wrist throughout the day, collects data such as the user's blood pressure, pulse, and orientation and transmits the collected data to the user's mobile phone or access point (AP) at home using a wireless ZigBee device. ZigBee is established by the ZigBee Alliance and adds network, security and application software to the IEEE 802.15.4 standard. Owing to its low power consumption and simple networking configuration, ZigBee is considered the most promising for wireless sensors.

Biomedical signals can be collected while moving in and out of the user's residence. The data collected inside of the house is sent to the AP in the house using Zigbee module. The AP stores the collected data and sends it regularly to the data storage at the healthcare center. When the user is outside of the house, the sensor sends the collected data to the user's mobile phone and then using CDMA module of the mobile phone, transmits the data to the center.

A light-weight data mining component is being developed for the mobiles and APs which briefly analyzes the data collected. This component has the responsibility of judging if an emergency occurs by analyzing the biomedical signals collected by the sensor. It also includes a function to call an emergency service using a motion detector attached to the sensor if it detects a fall-down, that is, when the user collapses.

3.2 Healthcare center

The healthcare center has two primary roles. First, it provides storage and management for the biomedical data collected from the users, and second, it monitors the users' health status and takes appropriate emergency or preventive action when required. A database server in the healthcare center stores and manages data including the medical, personal, family and other information for all registered users as well as biomedical signals collected from them. This data is used for real-time monitoring of users in case of emergencies and is also useful in periodic checkups.

The healthcare center also includes personnel who are stationed to keep monitoring users' health status and provide health information as well. Some of their responsibilities include regular phone checks, personal visits to users and emergency assistance if any abnormal signals are detected from a user.

3.3 CDSS (Clinical Decision Support System)

The CDSS supports long-term and short-term decision making processes by using models from distributed data mining, developing alternative plans and performing comparison analysis. In the short-term it assists in optimal planning to solve various decision making problems confronted in emergencies by utilizing the biomedical sig-nals. The goal of this system is to provide an information system environment where a decision maker can solve problems easily, accurately and promptly such that users are benefited. The CDSS needs to be integrated with a distributed data mining system that can provide global models.

3.4 Emergency response

Emergencies in a U-health framework require robust and quick recognition followed by an efficient emergency response. In this framework we employ a three pronged emergency recognition drive. Firstly, personnel monitoring the streaming biomedical data may detect abnormal signs and check user through phones or visits. Secondly, abnormal signs are also detected while mining the biomedical data collected over a period by the CDSS. Lastly, motion detectors mounted on sensors detect occurrence of falls and erratic movement.

The emergency management system uses a variety of hardware and software components that aim to improve emergency counteractions at the appropriate time and lower preventable deaths. This includes portable personal terminals comprising of RFID tags, portable RFID readers, an ambulance information system, a hospital in-formation system and a healthcare information system. The efficiency of the treatment in emergency rooms is increased by using RFID tags and readers. Since the system is well integrated it also transfers patient information in real-time to hospitals, and therefore medical teams who will provide treatment during emergencies can be well-prepared.

3.5 Short range wireless communication module

Biomedical signals collected from sensors are sent to mobile phones or APs using Zigbee, a short range wireless communication module. Zigbee is easy to control by complementing Bluetooth's weaknesses, provides multi hopping, and has low power consumption, which allows users to control the network size freely inside and outside of their houses (Hill et al., 2004). As Zigbee is a competitive short range wireless communication technology in vertical applications' area like a senor network, a large scale sensor network can be configured by combining a low power Zigbee transceiver and a sensor (Smithers & Hill, 1999).

3.6 Remote monitoring system

With increasing urbanization, shrinking of living space and shifting concepts of the family, elderly people often tend to live alone without any assistance at home. In such cases prompt responses are most important when a medical emergency occurs. The remote monitoring system is used to detect falls and erratic movement occurring at homes remotely using cameras or by checking current situations when an abnormal sign is detected. There may be signals that cannot be detected even with motion detectors mounted on sensors, or false alarms may occur. In these cases, the situations can be checked using in-house video cameras. The remote monitoring system is not only a management system for patient monitoring but aims for general health improvement of consumers through prevention of diseases, early detection, and prognosis management. Thus a customized personal healthcare service is established, maintained and controlled continuously (Jardine & Clough, 1999).

4. Clinical decision support with data mining

Data mining research is continually coming up with improved tools and methods to deal with distributed data. There are mainly two scenarios in distributed data mining (DDM): A database is naturally distributed geographically and data from all sites must be used to optimize results of data mining. A non-distributed database is too large to process on one machine due to processing and memory limits and must be broken up into smaller chunks that are sent to individual machines to be processed. In this paper we consider the latter scenario (Park & Kargupta, 2003). In this section we discuss how distributed data mining plays an important role within the CDSS component of the ubiquitous health-care system.

4.1 CDSS and DDM

In a ubiquitous healthcare framework DDM systems are required due to the large number of streams of data that have a very high data rate and are typically distributed. These need to be analyzed/mined in real-time to extract relevant information. Often such data come from wirelessly connected sources which have neither the computational resources to analyze them completely, nor enough bandwidth to transfer all the data to a central site for analysis. There is also another scenario where the data collected and stored at a center needs to be analyzed as a whole for creating the dynamic profiles. The preliminary empirical analysis with the prototype distributed data mining system discussed in this paper is suited towards this latter situation. The integration of the CDSS component of the ubiquitous healthcare framework with such a DDM is important.

As mentioned earlier the CDSS utilizes source data such as a user's blood pressure, pulse and temperature collected from the sensor, medical treatment history and other clinical data and integrates them for guidance on medical decision making. This involves both centralized and decentralized decision making processes and thus needs to employ distributed data modelling techniques. There are several levels of data mining involved in this process. Local mining of individual user data based on personalized medical history as well as global mining with respect to groups is required.

Data mining techniques used in the decision making system divide patients into groups. As a collection of patients have their own characteristics, they should be divided properly, and group properties are found through applying cluster analysis modelling techniques and searching created groups in the group analysis step. Secondly, finding causes and developing a model using mining techniques. Important causes of each subdivided group can be understood by the created cause and effect model, and through this, proper management for each patient can be achieved. Finally, a dynamic profile of the patient can be created using past history and domain knowledge in con-junction with sensory data. Each patient's risk rate is calculated by a system reflecting mining results, and administrators can see patients' risk rankings from the risk rates and give priority to patients with higher rates.

4.2 Distributed data mining architecture

This section describes a prototype system for DDM. For a detailed exposition of this system see (Viswanathan et al., 2000). The DDM system is build from various components as seen in figure 2. The DDM system takes source data and using SNOB (Wallace & Dowe, 2000), a mixture modeling tool, partitions it to clusters. The clusters get distributed over the LAN using MPI (developed by the Message Passing Interface Forum). Data models are developed for each cluster dataset using the classification algorithm C4.5 (Quinlan, 1993).

Finally the system uses a voting scheme to aggregate all the data models. The final global classification data model comprises of the top three rules for each class (where available). Note that MPI is used in conjunction with the known maximum number of hosts to classify the clusters in parallel using the C4.5 classification algorithm. If the number of clusters exceeds the available number of hosts then some hosts will classify multiple clusters (using MPI). Also the aggregation model scans all Rule files from all clusters and picks the best rules out of the union of all cluster rule sets. During the classification phase we have also classified the original dataset and produced rules modeling this data. To finally ascertain if our DDM system is efficient we compare our global model to this data model from the un-partitioned database. We compare the top three rules for each class from this model with our rules from the global model. If our global model is over 90% accurate in comparison to the data model from the original database we consider this as a useful result.

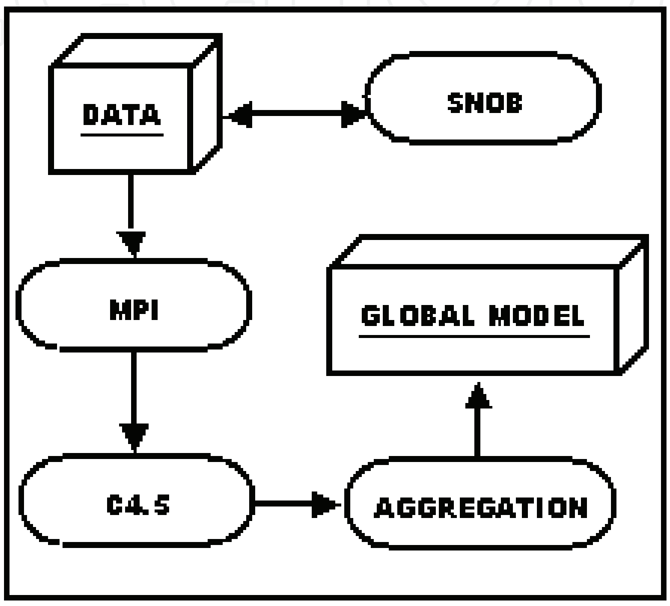


Fig. 2. DDM System Components

4.3 Preliminary results

The DDM system was tested on a number of real world datasets in order to test the effectiveness of data mining and the predictive accuracy. Detailed empirical analysis can be studied from (Viswanathan et al., 2005). In this section we present the DDM system performance results on the 'Pima-Indians-Diabetes' dataset from the UCI KDD Archive (Merz & Murphy, 1998). The diagnostic is whether the patient shows signs of diabetes according to World Health Organization criteria.

In order to study the usefulness of the system we compare the top three rules (where available) for each class from the partition-derived classification rules and rules from the original dataset. The aim of this testing is to find out the effect of our clustering process in partitioning, to the efficiency of our classification model and its predictive accuracy. We will consider 10% to be our threshold, average error rates of rules from partitions greater then 10% of that of the corresponding original rules is an undesirable result.

We can observe in figure 3 that the graphs comparing rules from partitions and original rules approximately follow the same gradient with the average error rate of partition rules staying above the original rules throughout with this gap closing as we approach higher classes. In general the distributed data mining system offers useful performance in the presence of a number of factors influencing the predictive accuracy. However many improvements and further research is needed in order to optimize the DDM system.

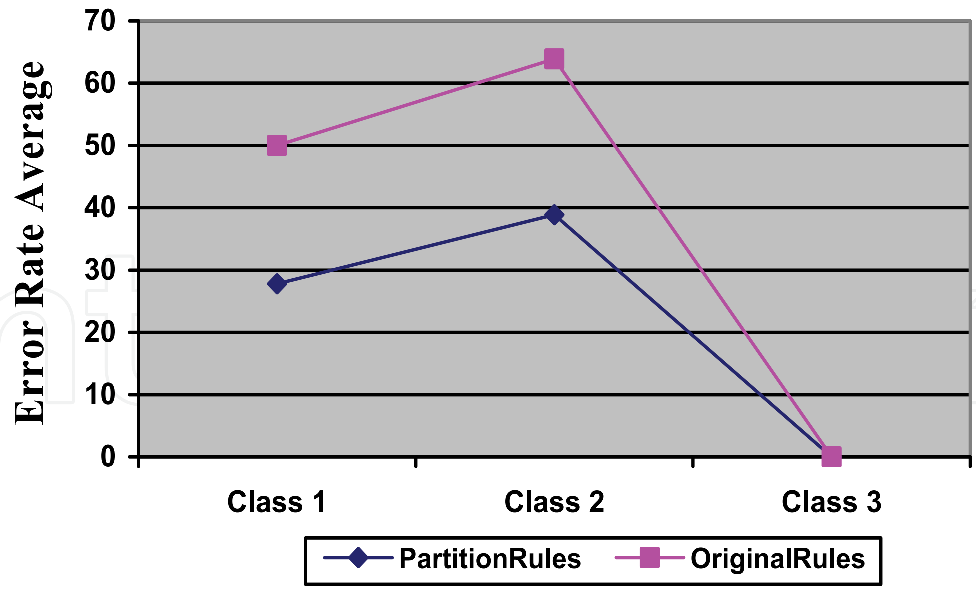


Fig. 3. Results from Partitioning

5. Conclusions and future challenges

As the elderly population constitutes a larger proportion of the aging society, providing quality long term care becomes an increasingly critical issue over the world. Our research aims to enable a patient-centric ubiquitous healthcare environment instead of the existing hospital-centric approach. The use of traditional verification-based approaches to analysis is difficult when the data is massive, highly dimensional, distributed, and uncertain. Innovative discovery-based approaches to health care data analysis with the integration of distributed data mining techniques warrant further attention.

This paper commences by describing a ubiquitous healthcare framework designed to provide consumers with freedom from temporal and spatial restrictions in their access to professional and personalized healthcare services anytime and anywhere – even outside of the hospital. Components of the system framework are discussed in brief. A prototype distributed data mining system is introduced with results from preliminary experiments on data. The plausibility of integrating such a DDM system with the clinical decision support component (CDSS) of the ubiquitous healthcare frameworks is highlighted.

However, there are several problems to solve, and the first one is accuracy. If sensors collect incorrect data, doctors can misjudge or misunderstand patients' emergency situations. Further analysis from the data mining mechanism is of great importance. The second is that there are controversial factors such as permissible ranges, certifications of doctors, and responsibility in case of the remote treatment. The existing law puts a limitation on the qualification of remote medical technicians, which impedes the spread of the system. Therefore, to activate the remote medical service, permissible ranges should be widened, and various remote medical technologies should be imported. The third is privacy protection. All user information employed such as bio-medical data collected from the remote monitoring systems or sensors should be handled with care to protect patients' privacy, and careful study is required to decide how much personal information should be open to the public. The fourth is security of biomedical signals. In this ubiquitous healthcare environment, sensors transmit collected biomedical signals to centers through wired or wireless communication, and these collected data are analyzed and used by the CDSS monitoring staff. Various security levels are required to control access to biomedical data stored in intermediate centers with access authorization.

6. References

CodeBlue. http://www.eecs.harvard.edu/~mdw/proj/codeblue.

EliteCare. http://www.elitecare.com/index.html.

Hill, J.; Horton M.; Kling R. & L. Krishnamurthy (2004). The Platforms enabling Wireless Sensor Networks. *Communications of the ACM*, Vol. 47 pp. 41-46.

Jardine, I. & Clough K. (1999). The Impact of Telemedicine and Telecare on Healthcare. *Journal of Telemedicine and Telecare*, Vol. 5, Supplement 1 127-128.

Kumar A. & Kantardzic M. (2006). Distributed Data Mining: Framework and Implementations, *IEEE Internet Computing*, vol. 10, no. 4, 2006, pp. 15-17.

Kristof Van Laerhoven et. al. (2004). Medical Healthcare Monitoring with Wearable and Implantable Sensors, *Proceedings of 2nd International Workshop on Ubiquitous Computing for Pervasive Healthcare Applications*.

MobileHealth. http://www.mobihealth.org/.

Message Passing Interface Forum (1994). "MPI: A message-passing interface standard". *International Journal of Supercomputer Applications*, 8(3/4):165-414.

Merz, C. & Murphy, P. (1998). *UCI repository of machine learning databases*. Irvine, CA: University of California Irvine, Department of Information and Computer Science. Internet: http://www.ics.uci.edu/~mlearn/MLRepository.html.

Park B. H., Kargupta H. (2003). "Distributed data mining: Algorithms, systems, and applications". *The Handbook of Data Mining*.Nong Ye (ed) Lawrence Erlbaum, New Jersey.

Quinlan, J. (1993). *C4.5: Programs for machine learning*. San Mateo, CA: Morgan Kaufmann.

Smithers C. R. and Hill N. Options for Wireless Technology in Telemedicine and Telecare Applications. *Journal of Telemedicine and Telecare*, Vol. 5, Supplement 1 138-139.

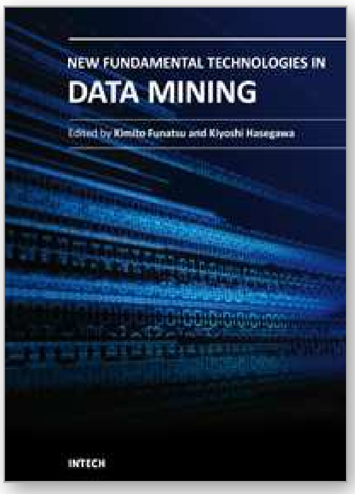
Smart Medical Home. University of Rochester, Center for Future Health, Rochester, NY U.S.A. http://www.futurehealth.rochester.edu/smart_home/Smart_home.html.

Tele-Monitoring: http://www.medical.philips.com.

United Nations Population Division Publications (2002). UN World Population Ageing 1950~2050. http://www.un.org/esa/population/publications/worldageing19502050/.

Viswanathan M.; Yang Y. K. & Whangbo T. K (2005). Distributed Data Mining on Clusters with Bayesian Mixture Modeling, *Lecture Notes in Computer Science*, Volume 3613, Jul, Pages 1207 – 1216.

Wallace C. & Dowe D. (2000). "MML clustering of multi-state, Poisson, von Mises circular and Gaussian distributions". *Statistics and Computing*, 10(1), pp. 73-83, January.



New Fundamental Technologies in Data Mining
Edited by Prof. Kimito Funatsu

ISBN 978-953-307-547-1
Hard cover, 584 pages
Publisher InTech
Published online 21, January, 2011
Published in print edition January, 2011

The progress of data mining technology and large public popularity establish a need for a comprehensive text on the subject. The series of books entitled by "Data Mining" address the need by presenting in-depth description of novel mining algorithms and many useful applications. In addition to understanding each section deeply, the two books present useful hints and strategies to solving problems in the following chapters. The contributing authors have highlighted many future research directions that will foster multi-disciplinary collaborations and hence will lead to significant development in the field of data mining.