Design of Intelligent Healthcare Support System Applying Machine Learning

MoonSun Shin, HaYoung Lee, JinSoo Kim* and JungHwan Kim

Dept. of Computer Engineering, Konkuk University, Republic of Korea
{msshin, hylee, jskim, jhkim}@kku.ac.kr

Abstract

As we move into an aging society, more and more people are interested in healthy old age and well-being. In this paper, we propose a personalized health information service system applying machine learning. The proposed personal health support system is focused on personalized recommendation service according to user profile of PHR(Personal Health Record). Machine learning algorithms such as the K-Nearest Neighbor algorithm and the data mining algorithms are adopted and applied to figure out characteristics of users and categorize the profiles of users. We design the proposed system that has an inference engine, which provides health contents or information for a specific user. The personalized healthcare service platform will make it possible to recommend personalized health information to the user. The proposed system can be used to handle smart self-care in an aging society.

Keywords: PHR, Personal Recommendation Service, Machine Learning, Data Mining

1. Introduction

Recently, medical services using ICT convergence technology and IoT technology have developed the medical industry in areas such as diagnosis, treatment and disease management. Healthcare services using ICT convergence technology and IoT technology not only cure diseases, but also activate diagnostic and management industries, while decreasing the proportion of traditional therapeutic industries, is leading to changes in market trends in disease treatment. The change of environment of medical service is evolving into 4P - oriented health care service for users and prevention and management according to the development of IT·BT technology[1][2]. The paradigm shift in health care services is shown n Figure 1.

This has the effect of reducing the cost of medical care for the entire country due to the reduction of the cost of treatment for the individual patient and the quality of the medical treatment is increased. Figure 1 shows the changes and needs in emerging medical environment. In 2016, global healthcare market was growing by 6.9% while health care IT market is expected to expand at an annual rate of 8.1% [3]. The main issues in the area of the healthcare IT are the development of home appliances connected with hospitals, virtual/remote monitoring, use of IT / BI solutions to reduce costs, and demand for profitability through regulation of healthcare data management systems[4]. There is a growing interest in the development of services that can help individuals to actively manage their health care in order to improve the quality of life and healthy old age.
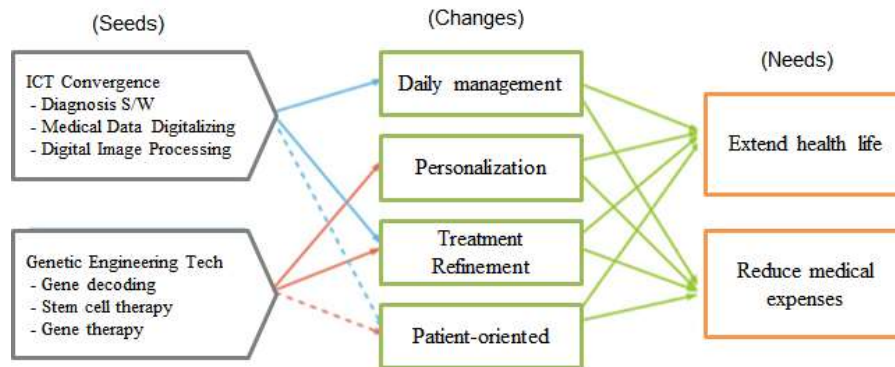
Therefore personalized intelligent smart health care is attracting more attention in an aging society. Using PHR(Personal Health Record) people can manage self-

healthcare through day life. In this paper, we design personalized healthcare support system based on machine learning for user-oriented health data management service.

Figure 1. Changes and Needs in Medical Environment

The user-oriented health data management service platform must provide the personalized healthcare support that matches the Health Care 3.0 era to meet the needs of the construction of smart devices for healthcare service utilizing social network[3]. The platform comprehensively manages to cover both PHR and life log data measured using smart device like sleep, emotion, exercise, food, and movement data[2]. Further, it can cover to provide appropriate health information according to user's PHR profile. To design personalized health information search algorithm, we apply the KNN(K-Nearest Neighboring) algorithm and the association rule mining algorithm, which were the useful machine learning techniques. The proposed system carries out four main functions such as manipulating PHR/life log data, analysis of user profile, recommendation of the proper health information for each user, and providing inference rules.

The rest of the paper is organized as follows. Section 2 describes the related research, and Section 3 presents the design of a personalized health care integration support system. In Section 4, we examine the analysis of user profile for intelligent smart healthcare service and design the algorithms for generation of similar user classification and reasoning rules. Section 5 describes the experiment and evaluation of the implemented algorithm. Finally, Section 6 provides the conclusion and future work.

2. Related Works

2.1. Ontology based Context-Aware

The context-aware system collects a lot of information from various sensors or devices, recognizes the user's situation through a specific algorithm, and provides necessary information to the user [6]. It is needed to construct a context aware model for accessing, managing, and presenting contextual information in the context-aware system. Methods for building context aware models include application-based approaches, model-based approaches, and ontology-based approaches [7]. The ontology was introduced to the field of computer science by S. H. Mealy in 1967 and recently it has been attracting attention as an important language expressing the semantic web in the field of artificial intelligence. The ontology is derived from the Greek words ontos (being) and logos (word), which were first used by German philosophers in the 19th century. It is defined as "providing a taxonomy that explains the kind of things that are present in the world, which are some aspects of the world, their nature and relationship characteristics".

Thomas Gruber defined ontology as "a formal and explicit specification of shared conceptualization" [8]. That is, the consensus knowledge that is not randomly set by some individual but is accepted by the consent of all the members involved in the area. The ontology is able to share the situation information and to understand and deduce the information because it can use the objective common format and give meaning to the information [9]. Each resource is connected with meaning, and rule-based reasoning is possible according to given rules. By using this point, it is possible to deduce information that is most suitable for the user's context. RDF, RDF Schema, and OWL are the most representative languages for expressing this meaningful information. Apache Jena, Pellet, and FaCT are platforms for helping to use this information [10].

In this paper, we propose ontology as a method of specification of health information modeling for personalized health information reasoning and recommendation for each user. By defining the health information specification applying the ontology, inference rules for health information search according to the context of user can be effectively used.

2.2. Machine Learning

Machine learning techniques are traditionally used for artificial intelligence or context awareness, and include various statistical and data mining techniques such as neural networks, Bayesian, Naive Bayesian, classification, decision trees, clustering, association rules, and sequential patterns. Data mining is a knowledge discovery technology that finds new data models or patterns that were not known from existing data to support making decision. In other words, discovering hidden relationships in data, data mining is a knowledge discovery technique that uses various techniques ranging from statistics to pattern recognition [11]. Representative data mining techniques include classification, clustering, and association rules. Classification identifies and classifies members within a particular group. Clustering creates similar groups based on the characteristics of the members. Association rule defines the relationship between concurrent events. Sequential pattern identifies the relationships between successive occurrences over a specific period of time, and forecasting is a technique that predicts future elements based on patterns in current data.

The association rule algorithm can search the association rules by setting the minimum support value from the random data and give meaning to the search result according to the confidence value. To make decision or recommendation service using association rule mining is meaningful under threshold like the support value and confidence value [11]. In this paper, we propose an attribute - based association rule algorithm.

The existing association rules algorithms cannot guarantee the efficiency of knowledge discovery for large amounts of data because of the problem of efficiency in generating all candidate sets. The proposed attribute - based association rules can improve the efficiency of existing methods by performing association rule exploration based on key attributes with the concept of weighting according to attributes [12]. This is useful for big data processing, and can improve performance and fast response time of recommendation and inference algorithms. In addition, by using the Ants Colony Decision Trees that are composed of the ants colony optimization and the decision tree, we generate the usefulness level of the health information contents thereby we use the collective intelligence such as page rank as the entropy value of the decision tree instead of pheromone of ants colony optimization [13][14].

3. Design of Intelligent Healthcare Support System

In this section, we describe the design of intelligent healthcare system which provides personalized health information to each user. First, we proposed an ontology-based contextual information model to implement PHR-based personalized health information recommendation service. Then we designed architecture of software and defined the main functions of reasoning.

3.1. Ontology based Context Aware Model for Healthcare Data

We defined PHR based healthcare data as ontology based contextual aware data model in order to analyze similar profile by applying inference algorithm and to search personalized recommendation information and to provide appropriate health information. We defined the classes, attributes, and constraints using OWL (Web Ontology Language). The concepts of each component of OWL were as follows. An ontology-based context aware data model using OWL for PHR-based healthcare system was shown in Figure 2.
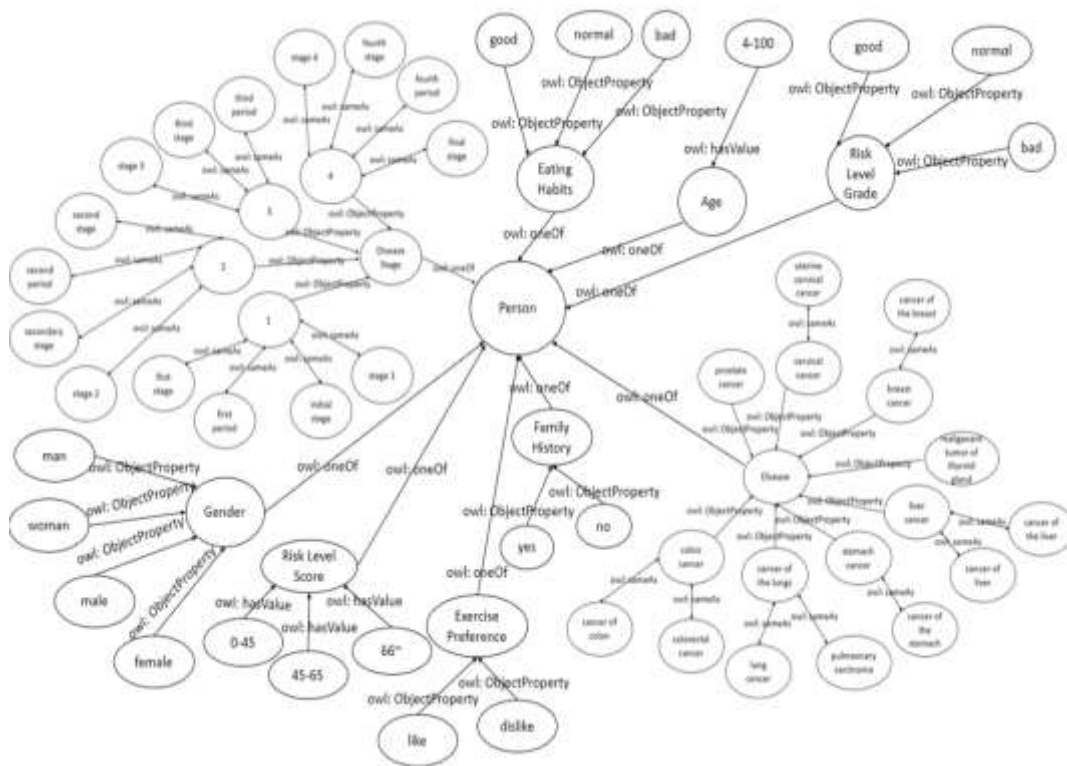

Figure 2. Ontology-based Context Aware Model using OWL

【Components of OWL】

- owl:Class Some concepts are abstracted and expressed. Every entity must belong to at least one owl: Class.
- owl:oneOf Enumerates instances of the OWL class. The domain of the owl: one of attribute is owl: Class, and the range is rdf: List.
- owl:ObjectProperty It is the class to which an object valued property belongs that relates an object to another entity.
- owl:hasValue For all instances of the subject Restriction class, it indicates that at least one of the constrained attribute values is equal to the value represented in the object.
- owl:sameAs   Indicates that two entities are identical even though the names are different. Both domain and range are defined by owl: a Thing.

【OWL Diagram】

The top-level class is defined as an object named Person, and the properties are enumerated as instances of the Person class. Attributes are sex, age, exercise preference, diet, family history, risk level, risk level, stage of illness, disease. Following Figure 3 showed OWL diagram for the personalized healthcare support system. Each attribute is oneOf, and the attribute value of gender, exercise preference, diet, family history, and risk grade is ObjectProperty.

The attribute value of age and risk level is hasValue. The property values below the disease stage and one stage of the disease become ObjectProperty, and the attribute values below them become the sameAs.

3.2. Design of Software Architecture

The personalized healthcare support system based on PHR proposed in this paper performs four services such as user profile analysis service, personal PHR and life-log record management service, daily life management according to PHR, and health content recommendation service according to user profile. Life-log data of each user and PHR information are stored in the database, and the clusters of similar users can be figured out by analyzing user's PHR profiles according to the individual health status. Then health information sharing service and the health related content recommendation service can be provided for each cluster [15].
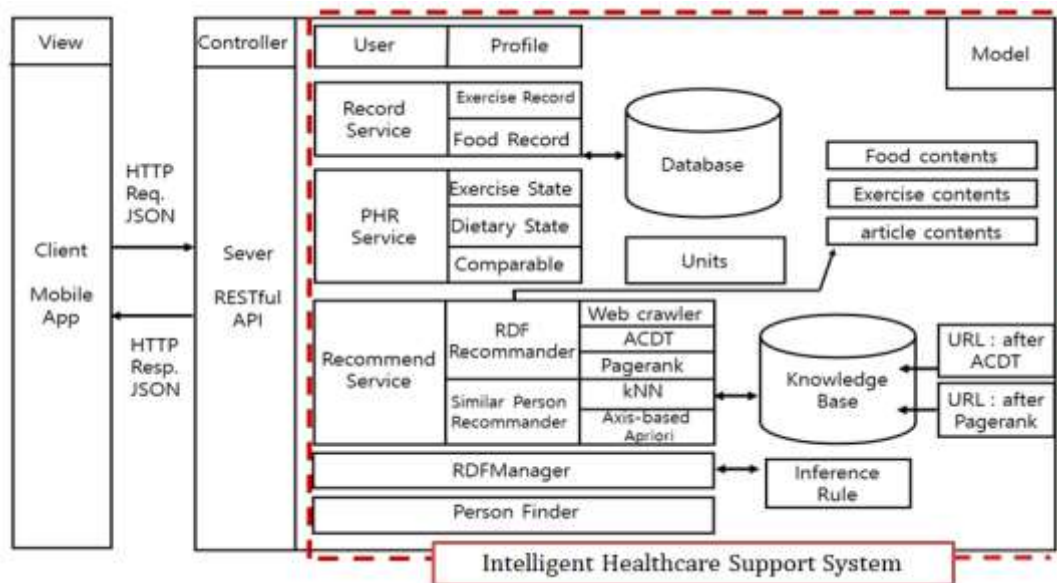
Figure 3. Architecture of PHR based Intelligent Healthcare Support System

The recommended service module is the core function of the intelligent health care support system. We applied and extended some of machine learning algorithms such as K-nearest neighbor and the association rule algorithm to design personalized recommendation information extraction algorithm. It is possible that these algorithms could be used to figure out the users' PHR profile, to classify similar state users, to compare similar biometric information and health information. At this time, the algorithm for discovering recommendation information uses an Ants Colony Decision Tree algorithm, which was motivated by the use of pheromones to find the optimal path when ants search for food [16]. Figure 3 showed the software architecture of the PHR-based customized health information recommendation service platform. In order to explore the inference rules, KNN algorithm was performed for similar profile search and then the

attribute based association rule mining algorithm was performed to generate association rules.

KNN algorithm could calculate the similarity of the user profile using Euclidean distance, and create the user group according to the similarity. The number of groups of similar users is adjustable and initial values are set so that 6 groups were created for the experiment. In order to analyze the user's profile and classify the health status, preprocessing of the PHR data should be performed so that the data preprocessing process is required and the scores are classified according to the six groups and classified according to the scores. In order to analyze the user's profile and classify the health status, preprocessing of the PHR data should be carried out. So we performed the data preprocessing process and the data values were changed to discrete values according to the six groups.


(a) KNN   (b) Attribute-based Association Rules

Figure 4. Algorithms for Reasoning in Intelligent Healthcare Support System

The pseudo code of KNN algorithm and the attribute-based association rule algorithm proposed in this paper were shown in Figure 4. The proposed attribute-based association rule algorithm was designed to extend the existing Apriori algorithm, which has been a typical association rule algorithm, so first to select important attributes and then perform association rules based on the selected attributes. This could increase the efficiency of knowledge discovery by filtering the generation of candidate item sets for unnecessary or unimportant attributes.

After probing the inference rules generated by similar user classification and association rules, the recommend engine performed the internet crawl using the web crawler in order to explore user-customized health related contents for personalized recommendation service. For the optimization of the personalized recommendation service ACDT(Ants Colony Decision Tree) was applied because ACDT could provide incremental rank of the contents. The sequence diagram for intelligent healthcare support system was shown in Figure 5.
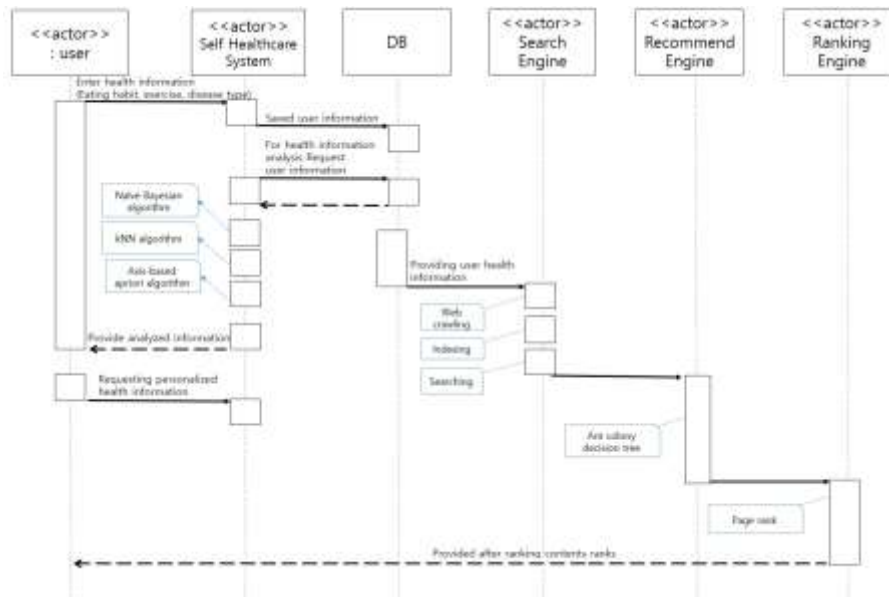
Figure 5. Sequence Diagram for Intelligent Healthcare Support System

After finding antecedent information using an ACDT, which was modified by an ants colony optimization algorithm, the page rank would be used to expose more important data to the upper part, thereby providing optimized health information contents to users. ACDT algorithm was shown in Figure 6.

Figure 6. Ants Colony Decision Tree Algorithm

The ACDT algorithm was a combination of existing ants colony optimization algorithm and decision tree [16]. Equation (1) represented the optimization equation of the ants colony optimization algorithm.

(1)

$$\tau_{ij} = (1 - \rho) \times \tau_{ij} + \sum_{k=1}^{m} \Delta\tau_{ij}^{k}$$

$$\Delta\tau_{ij}^{k} = \begin{cases} \frac{Q}{L_k} & \text{when } k \text{ ant move vertex } i \text{ to vertex } j \\ 0 & \text{in other cases} \end{cases}$$

It was not suitable for decision tree because it is optimized for general ants colony optimization algorithm. Therefore, to make an equation applicable to decision tree, it must be transformed into the following equation (2).

$$\tau_{ij}(t+1) = (1-\rho) \times \tau_{ij}(t) + \rho \times E \qquad (2)$$

E in equation (2) was an entropy calculation formula of decision tree. It was appropriate to apply the decision tree algorithm to obtain the pheromone value. Entropy, E, must be calculated as follows in equation (3).

$$E(T) = -\sum_{i=1}^{k} p_i \log_2 p_i \qquad (3)$$

In equation (3), p was the category ratio, and the data set T was divided into k categories and the category ratios are the same values as $p_1, p_2, \cdots, p_k$. In this case, the categories were generated based on the content classification and the user's self-diagnosis questionnaire list, and the category ratios were determined by themselves. Therefore equation (2) was a result of multiplying the evaporation rate by entropy E and increasing the evaporation rate as the content impurity increased, thereby updating the pheromone value. This was applied to the Web crawler and the page rank algorithm. Before the ACDT algorithm was executed, the keywords used in the crawl had been selected by referencing the ontology-based health information model and the URL had been crawled by the web crawler and stored in the databases. Thereafter, URLs mixed together indiscriminately were filtered and classified only for valid information according to the health information of the user by applying the ACDT. At this time, the classified URLs were stored in the knowledge base. The PageRank algorithm was categorized by an ACDT, and the URLs stored in the knowledge base were sorted order by importance and displayed to the user. Through this process, a lot of redundant information could be removed, and when the pheromone value would become zero, the path was not used at all, so that unnecessary information could be filtered. Also, massive information could be classified and information could be organized. Therefore, it was expected that the ACDT could be useful for the recommendation service because it would be able to provide optimized customized information to the users.
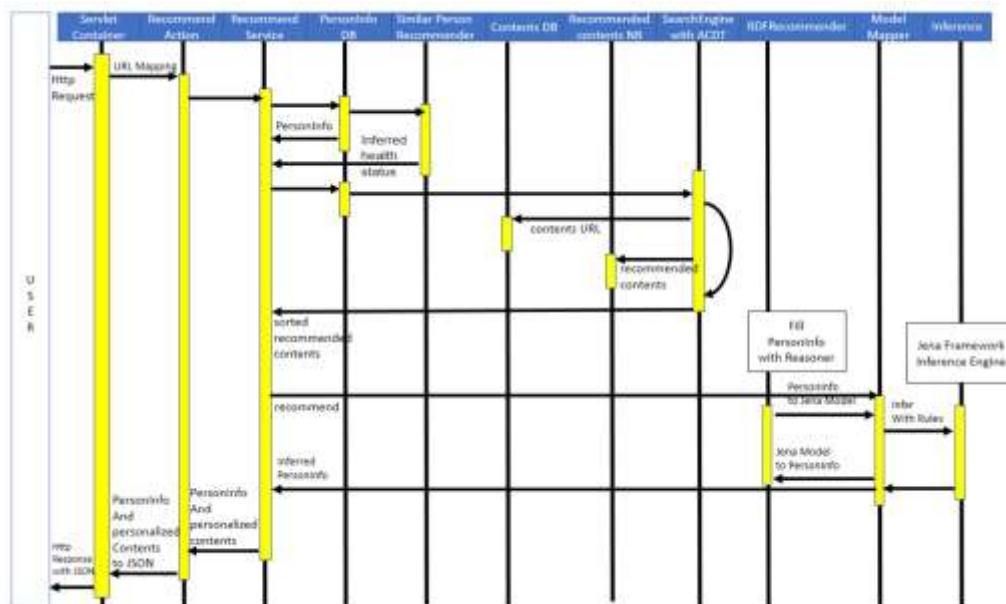
Figure 7. Inference Workflow in Intelligent Healthcare Support System

These algorithms were applied to the basic workflow as shown in Figure 7. The process of user context awareness, profile analysis, content recommendation and

reasoning system had used these algorithms in order to infer personalized health information. And inference rules have been incrementally stored in the knowledge base. The KNN algorithm and the attribute-based association rule algorithm were used to support the module of the Similar Person Recommender, which performed to infer the user's health state. ACDT algorithm was included in the module named Search Engine with ACDT, where the crawler collected the URLs of health information contents which would be stored in the databases, Then ACDT was executed on this database for the contents optimization. The customized health information URL, which was the execution result of ACDT algorithm, was stored in the knowledge base called Recommended Contents, and the knowledge base would be updated after the PageRank was executed. This process is repeated so that the knowledge base in which it could be possible to maintain optimized contents for users. Figure 7 showed the workflow of the inference process in the intelligent healthcare system.

4. Experimental Results
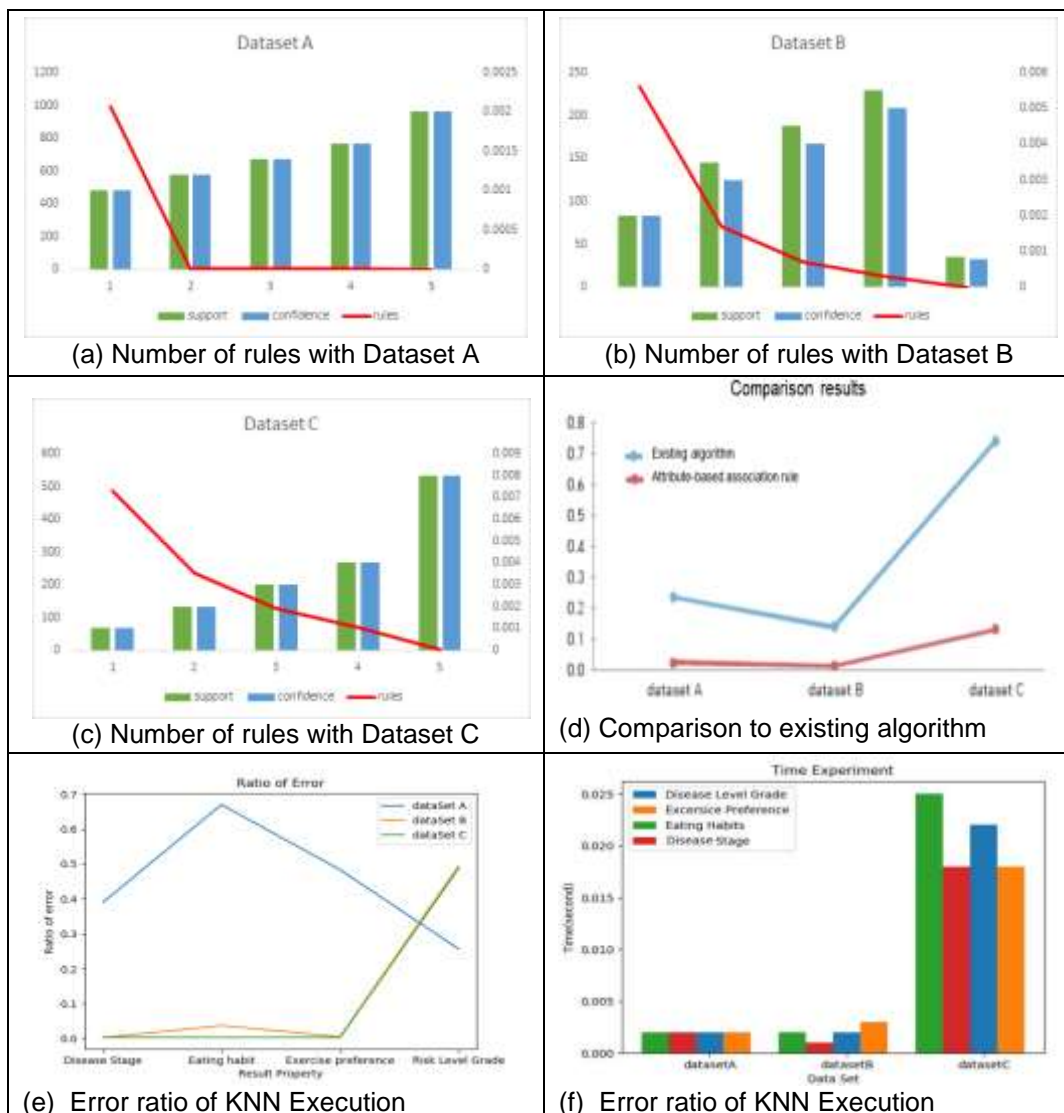
4.1. Experimental Data Preprocessing

First, we defined the data requirement specification for intelligent health care support system for health information modeling and evaluation of machine learning algorithms. Experimental data was generated by random sampling from the general health data. Experiments were carried out to measure algorithm performance. The test data consisted of general numerical data and range values, so data preprocessing was needed. And data values such as dietary habits, exercise preference, family history, and disease stage were converted into discrete values and the experiment and evaluation were conducted. Three dataset were prepared for the experiments. Dataset A, an arbitrary unqualified dataset, consisted of 1000 tuples. Dataset B consisted of 1000 tuples with some patterns and preprocessing. For the comparison of the proposed algorithms to existing ones, dataset C was prepared with 10000 tuples.

Experimental data are classified into sex, age, exercise preference, diet, disease, disease state, and family history. Exercise preference is an indicator of whether a user likes or dislikes exercise, and eating habits are divided into good, normal, and bad. In this experimental data, the disease was limited to cancer such as lung cancer, stomach cancer, liver cancer, thyroid cancer, breast cancer, cervical cancer and prostate cancer. The progression ranks of these cancers were handled into the disease state. Family history was confirmed by the presence of family members who had cancer among their immediate family members. For the evaluation of the efficiency the data sets were also distinguished having rules or no rules. The first preprocessing was data conversion to perform association rules, and the result was numerical data and character data conversion. The second preprocessing was carried out to convert the numerical data, and as a result, it became possible to execute KNN in order to classify similar user groups for recommendation of customized health contents. The primary preprocessing attributes of a dataset were: The gender was divided into male and female. Age was between 14 and 100 years of age. The exercise preference was divided into like and dislike. The eating habits were classified as good, normal, and bad. In addition, the type of cancer was selected based on the incidence rate of cancer in the National Cancer Information Center. Disease stages are divided into 1, 2, 3, and 4 groups. Family history was divided into 1 (yes) and 0 (no).

4.2. Experimental Results

In order to carry out experiments on various experimental data sets, the number of experimental data was changed. Also we figured out the patient's exercise preference and eating habit according to the disease stage, so that customized contents can be applied to

the recommendation service providing useful information. Experiments have shown that it is possible to recommend the contents based on the patient's condition according to the state of illness and disease, and to provide an efficient search time for recommendable information exploration. The test of attribute-based association rules were carried out based on the specific attributes like exercise preference, eating habits, and disease stage attributes, which were considered to be important attributes. This experiment was performed to determine optimal value of minimum support and confidence for each dataset and to verify that the algorithm worked. The graphical results of these experiments were shown in Figure 8(a), Figure 8(b), and Figure 8(c). In addition, we can confirm that one attribute can be derived by referencing two or more attributes, so that it would be possible to deduce missing information even if all the health information about the user could not be provided.

(a) Number of rules with Dataset A

(b) Number of rules with Dataset B

(c) Number of rules with Dataset C

(d) Comparison to existing algorithm

(e) Error ratio of KNN Execution

(f) Error ratio of KNN Execution

Figure 8. Graphs of Experimental Results

The result of comparison test of the existing association rule algorithm and the property-based extended association rule algorithm for each test data set was shown in Figure 8 (d). In particular, it can be seen that attribute-based association rules show excellent performance regardless of the number of datasets B and C.

We analyzed the performance and error rate of the KNN algorithm when the final inferred results for the second pre-processed dataset A, B, and C. The Horatio of KNN algorithm was set as 0.3. This experiment was performed to figure out the best efficiency of the final results of inference. As it can be seen in Figure 8(e), the error rates of the regular data sets B and C are remarkably low. The error rates of most attribute that were selected to be important were low except attribute risk level grades. Therefore, it means that similarity of data is meaningful to execute KNN on the attribute such as exercise preference, dietary habits, and disease states. Figure 8(f) showed that it took about 0.0025 seconds for about 1,000 data and about 0.02 second to 0.025 second for over 10,000 data regardless of the data type. Therefore, even if the amount of data increases, similar user information must be handled effectively by the intelligent healthcare system we proposed.

5. Conclusion and Future Works

In this paper, we proposed PHR-based intelligent healthcare support system applying machine learning. We defined ontology-based health data context model, where the ontology was used as a way of defining the specification of the user's health information and to utilize the inference rules for the health information search according to the situation. Also we designed software architecture of the system and extended some algorithms such as KNN, attribute–based association rules and ACDT in order to provide users with optimized personalized health information. We have verified the usefulness of the proposed system through the experiments of the algorithms we applied and extended.

We are going to develop the prototype of the PHR-based healthcare support system. We have constructed healthcare server system and keep going to develop smart phone app as a client program. Our works will be contributed to not only for intelligent healthcare management but also for better life in the coming aging society.

Acknowledgments

References

[1]   S. H. Kim, "Health-IT convergence technology trends and forecasts (mainly U-health)", Journal of Electronics, vol. 37, no. 6, **(2010)**.
[2]   M. S. Shin, H. S. Jeon and B. J. Lee, "Constructing RBAC based Security Model in u-Health Care Service Platform", TSWJ (937914), **(2014)**.
[3]   J.Y. Choi, "Healthcare 3.0: A world of healthy life", CEO information, Samsung Economic Research Institute, vol.831, **(2011)**.
[4]   K. T. Oh and J. E. Lee, "Smart Life revolutionary real and smartphone addiction", Internet and Information Security, vol.3, no.4, **(2012)**.
[5]   J. H. Park and T. K. Hwangbo, "Healthcare IT Convergence Technology", Journal of Korea Information and Communications", vol.28, no 5, **(2011)**.
[6]   M. S. Shin, M. C. Ko, Y. W. Ju, Y. J. Jung and B. J. Lee, "Implementation of Context-Aware Based Robot Control System for Automatic Postal Logistics", Studies in Informatics and Control, ISSN 1220-1766, vol. 22, no. 1, **(2013)**, pp. 71-80.
[7]   Ledey, A.G. Abowd, "Towards a Better Understanding of Context and Context-Awareness", Workshop on the what, who, where, when and how of context-awareness at CHI 2000, **(2000)**.
[8]   D. Stevenson, "Requirements and Design for Smart Healthcare Applications", EJHI2011, vol.6, no.2, **(2011)**.

[9]   W3C, "OWL2 Web Ontology Language", W3C Recommendation 27 October, 2009, http://www.w3.org/TR/owl12-overview/.

[10]   H. Chung and J. Kim, "Design of Semantic Models for Teaching and Learning based on Convergence of Ontology Technology", Journal of the Korea Convergence Society, vol. 6, no. 3, **(2015)**, pp. 127-134.

[11]   Y. Sung Cho and S. Chul Moon, "Weighted Mining Frequent Pattern based Customer's RFM Score for Personalized u-Commerce Recommendation System", Journal of Convergence, vol.4. no.4, **(2013)**.

[12]   M. Brahami, B. Atmani and N. Matta, "Dynamic knowledge mapping guided by data mining:Application on Healthcare", Journal of Information Processing Systems, vol.9, no.1, **(2013)**.

[13]   J.S. Sohn, U.B. Bae and I.J. Chung, "Contents Recommendation Method Using Social Network Analysis", Wireless Personal Communications, vol. 73, **(2013)**, pp. 1529-1546.

[14]   M. Nilashi, O. B. Ibrahim and N. Ithnin, "Hybrid recommendation approaches for multi-criteria collaborative filtering, Expert Systems with Applications, vol. 41, **(2014)**, pp. 3879-3900.

[15]   M. Dorigo and C. Blum, "Ant colony optimization theory: A survey", Theoretical computer science, vol. 344, **(2005)**, pp. 243-278.

[16]   U. Boryczka and J. Kozak, "Ant colony decision trees–A new method for constructing decision trees based on ant colony optimization", in Computational Collective Intelligence. Technologies and Applications, Springer, **(2010)**, pp. 373-382.

Authors

Moonsun Shin, she received Ph.D degrees from Chungbuk National University in 2004 in computer science. In 2005 she joined the faculty of Konkuk University, where she is now an associate professor. Her research interests include context awareness, security model. ICT convergence and big data analysis.

Hayoung Lee, she is an undergraduate student in Computer Engineering at KonKuk University. She is going to study as a graduate school for M.S at Queensland University of Technology, Australia. Her research interests include ontology-based data model, ICT convergence and big data analysis.

Jinsoo Kim, he received the B.S. degree from Seoul National University in 1983, and the M.S. and Ph.D degrees from Korea Advanced Institute of Science and Technology (KAIST), in 1985 and 1998. In 1985, he joined Korea Telecom, where he was a senior researcher. In 2000, he joined the faculty of Konkuk University, where he is now a professor. His research interests include Internet packet processing, high-speed networks, information centric networks, and wireless sensor networks.

Junghwan Kim, he received the B.S., M.S. and Ph.D degrees from Seoul National University, in 1991, 1993 and 1999, respectively, all in computer science. In 1999 he joined Samsung Electronics, where he was a senior researcher. In 2001 he joined the faculty of Konkuk University, where he is now a professor. His research interests are in the areas of parallel computing, communication networking, GPU computing, and design of efficient algorithms.