International Journal of Bio-Science and Bio-Technology
Vol.8, No.2 (2016), pp.227-236
http://dx.doi.org/10.14257/ijbsbt.2016.8.2.21

Developing an Intelligent Health Pre-Diagnosis System for Korean Traditional Medicine Public User

Kwang Baek Kim[1*] and Doo Heon Song[2]

[1]Department of Computer Engineering, Silla University, Busan 617-736, Korea
[2]Department of Computer Games, Songdam College, Yongin 449-040, Korea
[1]gbkim@silla.ac.kr, [2]dsong@ysc.ac.kr

Abstract

Korea Traditional Medicine is popular in use among Korean public but there exist few available health information systems on the internet. A computerized self-checking diagnosis system is proposed to reduce the social cost by monitoring health status with simple symptom checking procedures especially for Korea Traditional Medicine users. Methods: Based on the national reports for disease/symptoms of Korea Traditional Medicine, we build a reliable database and devise an intelligent inference engine using Fuzzy ART Clustering. The implemented system gives five most probable diseases a user might have with respect to symptoms given by the user. Thus, in this paper, we propose a self-diagnosis system of Korean traditional medicine based on Korean Standard Causes of Death Disease Classification Index(KCD) and Fuzzy ART/inference method. Since this is for self-diagnosis, our system has graphical user-friendly interface that accepts symptoms of user from a certain part of body where the user feels inconvenient. Inference results are verified by Korea Traditional Medicine doctors as sufficiently accurate and easy to use.

***Keywords**: Health Diagnosis, Korean Traditional Medicine, Fuzzy ART, Inference*

1. Introduction

Computer-assisted medical diagnosis system has long history of research and requires analyzing bulky test cases, gathering field experts' opinions, and uncertainty management algorithms. Usually such diagnostic systems are designed for medical doctors giving more accurate decision making and therefore the diagnosis/treatment consultation is limited to a certain set of diseases or body parts with deep knowledge [1, 2].

However, diagnosis based on Korean traditional medicine is not easy to understand by the public as its inference mechanism is largely metaphorical or abstract. While there are plenty of internet services for westernized medicine for self-diagnosis, causes and treatment information, there are few such services for Korean traditional medicine. Furthermore, those rare services can only give information of treatments and symptoms with given name of disease. Thus, it is post-hoc supplementary information after expert's diagnosis.

The main difficulties in building informative pre-diagnosis or self-diagnosis system is that it requires a reliable symptom-disease database and disease classification system. For the symptom-disease database, since Korean traditional medicine has been built upon the innate characteristics of Korean people's body [3-6], we need such specialized evidences for Koreans and happily the government (Statics Korea, http://www.kostat.go.kr) has published Korean Standard Causes of Death Disease Classification Index(KCD) that can be a perfect starting index of such Korean specialized database build-up.

For the disease classification system, we need an unsupervised learning method since a supervised learner such as neural network algorithm causes frequent relearning as

*Corresponding Author: Kwang Baek Kim

ISSN: 2233-7849 IJBSBT

symptoms and diseases are to be added. Thus in this paper, we propose a self-diagnosis system for Korean traditional medicine with a disease database based on KCD and a fuzzy ART algorithm in conjunction with fuzzy inference engine as a classification system. Previous studies [7-11] have explored several neural/fuzzy unsupervised learning algorithms and found that fuzzy ART is most appropriate with its stability and accuracy in diagnosis.

Given symptoms from the user, guided by graphical user interface, our system extracts five most probable diseases with causes and treatments. In those processes, fuzzy inference algorithm plays a critical role to pick up target diseases accurately. The performance of our system as well as relationships among symptoms and diseases and their causes/treatments are verified by experts in Korean traditional medicine and the result is sufficiently competitive as a self-diagnosis system.

2. Preparing Database

2.1. Data Collection

We collect 739 diseases and 363 related symptoms based on KCD (Korean Standard Causes of Death Disease Classification Index) which replaces diseases of ICD (International Causes of Death) published by WHO with Korean traditional medicine. Information on causes and treatments of such diseases are extracted from many textbooks of Korean traditional medicine including well known folk remedies. Collected information is verified by doctors in Korean traditional medicine three times to reduce conflicts and errors.

Figure 1 shows a snapshot of relationships between symptoms and diseases in our database. Collected symptoms are classified into 17 groups with respect to related human body parts. Figure 2 summarizes database construction processes.

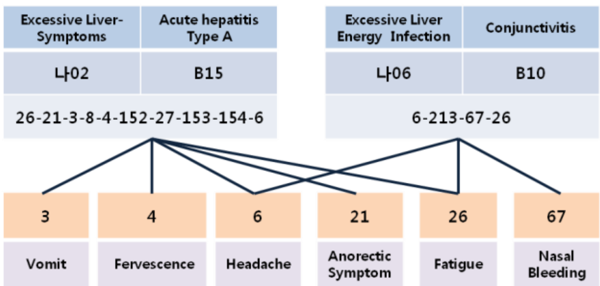

Figure 1. Relationships between Symptoms and Diseases

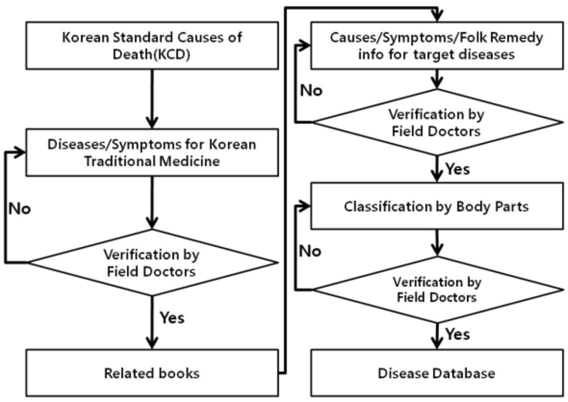

Figure 2. Process of Constructing Disease Database Collection and Symptoms with Respect to Related Body Part

2.2. Database Entity Structure

Database entities are designed as shown in Table 1. It includes user information and disease information. Disease information includes ICD code and KCD code and names of diseases in Korean and western medicine so that user can get maximal information about the disease.

Table 1. Entity Definitions

Entity	Explanation
disease	Disease code, ICD code, KCD code, Disease names(Korean medicine and Western Medicine), Causes, Treatments
symptom	Symptom code, Symptom name, related body part
bodypart	Body part code, Part name
guest	User information
doctor	Expert/medical doctor's information

3. Disease Classification System with Fuzzy ART and Fuzzy Inference Algorithm

In general, neural network learning algorithms have common inefficiency in that they require frequent relearning steps as diseases/symptoms are added/deleted/updated since they are supervised learners. Thus, we need a unsupervised learner that has maximal independencies on clusters with respect to input. Previous studies [7-9, 11] tested various forms of FCM (Fuzzy C-means) and ART algorithms in diagnosis and verified that an enhanced Fuzzy-ART is the best in stability.

In that algorithm [11], with given user input (most representative symptoms), it sends queries to database to suppress selecting unrelated symptoms. However, for some diseases that have different symptoms in their early and later stage, or disease like flu which has too many symptoms but some are inherent to user, this algorithm has a certain level of weakness because in those cases, by the characteristics of the Fuzzy ART that applies average of input patterns in controlling weights, the similarity between disease vector and symptom vector becomes unexpectedly low. Thus, we add a fuzzy inference engine to mitigate that weakness. The overall process of disease classification system is described as Figure 3 and explanations at each step will be followed.

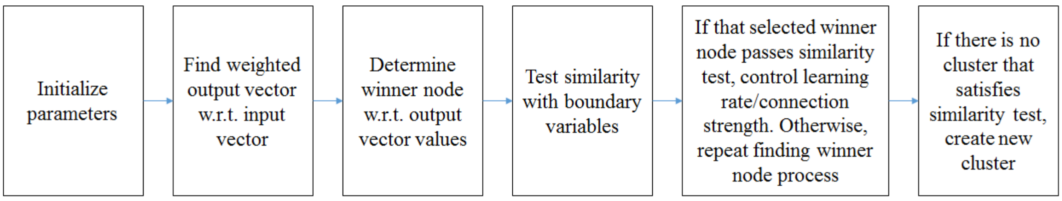

Figure 3. Processes of Fuzzy ART algorithm in Disease Classification

In step 2, the weight is to set the maximum of output vector as 1. There are cases that the maximum of output vector is less than 1 due to the inconsistency between input binary vector and connecting weight vector that has continuous nature. This weight is applied to the disease classification and membership function of disease with respect to symptoms in the learning process.

In step 3, there are two different cases for computing output vector. The first case is for the early learning process. The output vector is the minimum of fuzzy membership value divided by output weight and boundary variable value divided by output weight. The winning neuron is determined by taking the maximum of such output vector values.

In that case, the process continues to step 4 to step 6 such that after investigating the similarity between boundary variables, it determines if a new cluster is required. If it is sufficiently similar (step 5), it controls learning rate and connection strengths. Otherwise, it creates a new cluster (step 6). The final output vector means the degree of membership belongs to the disease. The flow chart of such process is shown in Figure 4 [9].

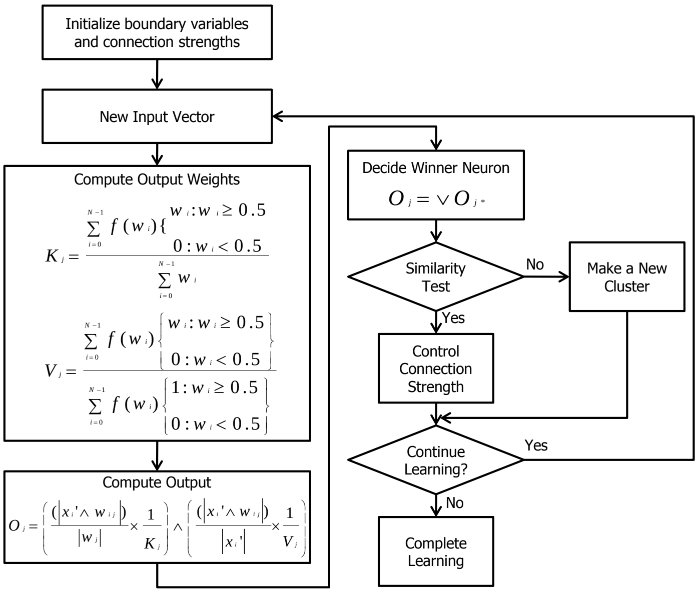

Figure 4. Flow Chart of the First Case

The second case is when a user wants to know the diagnosis with input. The criteria to determine five most probable diseases are related to the agreement rate between user symptoms and disease symptoms (variable X) and the number of symptoms belongs to the disease (variable Y). Values of two variables are computed by related membership functions and then fuzzy inference rules are applied to finalize the result. Then, it is multiplied to the value of the first case to control the degree of membership with respect to user input.

$$O_j = \sum_{i}^{n-1} Output[i] * Ot \tag{1}$$

In formula (1), *Output* is the output of the first case and O_t is the result of the second case after applying fuzzy inference rules. If O_t is high, *Output* also becomes high because high O_t means that it is highly related to the user input symptoms and if O_t is low, by the same reason, *Output* also becomes low. O_t is computed as formula (2).

$$O_t = Fuzzy\ (\#\ of\ symptoms\ according\ with\ user\ symptoms,\ \#\ of\ symptoms\ of\ all\ diseases) \tag{2}$$

There are three fuzzy intervals with respect to the agreement rate of symptoms between user symptoms and disease symptoms, L(low), M(medium), H(high) and its membership function is shown as Figure 5.

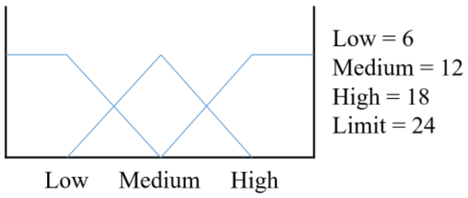

Figure 5. Membership function for the agreement rate between user symptoms and disease symptoms

When the agreement rate between user symptoms and disease symptoms is low

$$
\begin{aligned}
&If\ (X \le low)\ then\ \mu(X) = 1 \\
&Else\ If\ (X \ge medium)\ then\ \mu(X) = 0 \\
&Else\ \mu(X) = \frac{medium - X}{medium - low}
\end{aligned} \tag{3}
$$

When the agreement rate between user symptoms and disease symptoms is medium

$$
\begin{aligned}
&If\ (X \le low)\ or\ (X \ge high)\ then\ \mu(X) = 1 \\
&Else\ If\ (X \ge medium)\ then\ \mu(X) = \frac{high - X}{high - medium} \\
&Else\ If\ (X \le medium)\ then\ \mu(X) = \frac{X - low}{medium - low}
\end{aligned} \tag{4}
$$

When the agreement rate between user symptoms and disease symptoms is high

$$
\begin{aligned}
&If\ (X \ge high)\ then\ \mu(X) = 1 \\
&Else\ If\ (X \le medium)\ then\ \mu(X) = 0 \\
&Else\ \mu(X) = \frac{X - medium}{high - medium}
\end{aligned} \tag{5}
$$

There are three fuzzy intervals with respect to the total number of symptoms belong to the disease, L(low), M(medium), H(high) and its membership function is shown as Figure 6.

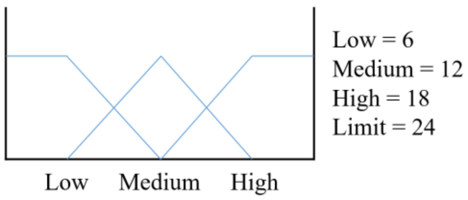

Figure 6. Membership Functions for the Number of Symptoms belongs to the Disease

When the total number of symptoms belongs to the disease is low

$$
\begin{aligned}
&If\ (Y \le low)\ then\ \mu(Y) = 1 \\
&Else\ If\ (Y \ge medium)\ then\ \mu(Y) = 0 \\
&Else\ \mu(Y) = \frac{medium - Y}{medium - low}
\end{aligned} \tag{6}
$$

When the total number of symptoms belongs to the disease is medium

$$
\begin{aligned}
&If\ (Y \le low)\ or\ (Y \ge high)\ then\ \mu(Y) = 1 \\
&Else\ If\ (Y \ge medium)\ then\ \mu(Y) = \frac{high - Y}{high - medium} \\
&Else\ If\ (Y \le medium)\ then\ \mu(Y) = \frac{Y - low}{medium - low}
\end{aligned} \tag{7}
$$

When the total number of symptoms belongs to the disease is high

$$
\begin{aligned}
&If\ (Y \ge high)\ then\ \mu(Y) = 1 \\
&Else\ If\ (Y \le medium)\ then\ \mu(Y) = 0 \\
&Else\ \mu(Y) = \frac{Y - medium}{high - medium}
\end{aligned} \tag{8}
$$

Then the disease membership function O_t is shown as figure 7.

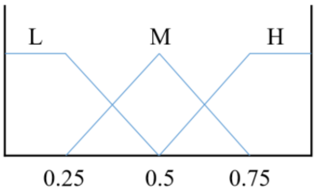

Figure 7. Disease Membership Functions for the User Input

The "if-then" style fuzzy inference rules (Figure 8) to determine most probable disease based on user input are as following by combining above two fuzzy membership functions. We apply Min-Max inference method in computing membership degree of diseases. The overall flow of the second case can be summarized as Figure 9.

R1 : If X is L, Y is L then α is L
R2 : If X is L, Y is M then α is M
R3 : If X is L, Y is H then α is M
R4 : If X is M, Y is L then α is L
R5 : If X is M, Y is M then α is M
R6 : If X is M, Y is H then α is H
R7 : If X is H, Y is L then α is M
R8 : If X is H, Y is M then α is H
R9 : If X is H, Y is H then α is H

Figure 8. Fuzzy Inference Rules

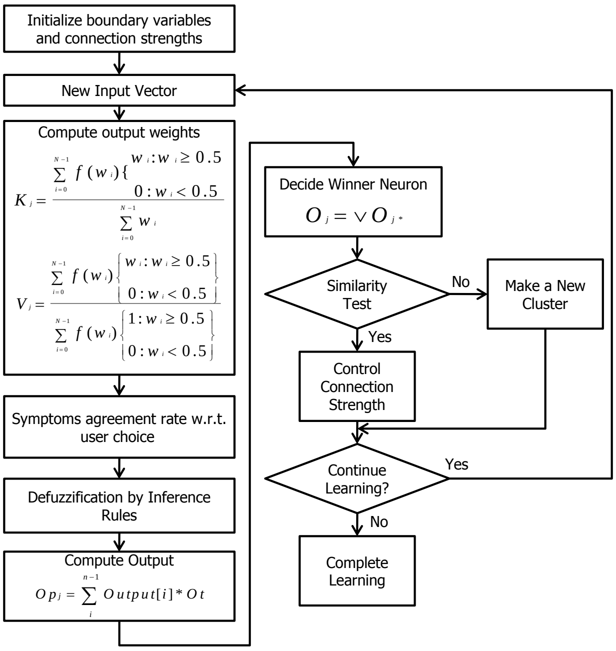

Figure 9. Flow Chart of the Second Case

4. Experiment and Analysis

The implementation environment is as following; a IBM compatible PC with Intel Pentium IV 2 GHz CPU and 1G RAM is used and Eclipse 3.2, Apache Tomcat 5.5, Apache 2.2, Adobe Photoshop 7.0, JDK 1.6 and Oracle 10g are used in implementation and the system is available for on-line environment using JSP. Figure 10 shows the block diagram of the proposed system.

Users are guided by the graphical interface shown as Figure 11 to give input symptoms. Symptoms are classified into 17 representative body parts and the user clicks the body part from the human body image to select the input symptoms. Then, the system outputs five most probable diseases with their causes and treatments by Fuzzy ART algorithm and fuzzy inference rules explained in section 3 as shown in Figure 12.

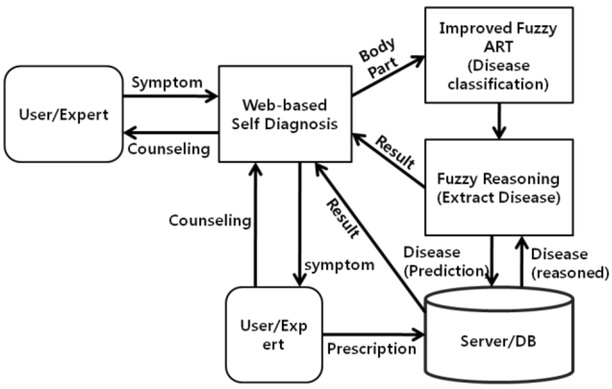

Figure 10. Block Diagram of the Proposed Self-Diagnosis System

Figure 11. Selecting User Symptoms from the Proposed System

Figure 12. Screen for Extracting the Disease

This result can be compared with the result of previous study [9] that uses Fuzzy ART but no fuzzy inference rules. In order to have fair comparison, we give the same input – cephalalgia (headache), fatigue, and anorexia to systems with and without fuzzy inference rules. When fuzzy inference rules are not applied, five most probable diseases are extracted as shown in Figure 13 and the number of agreed symptoms for those diseases is shown as Table 2.

Figure 13. Result Screen by Fuzzy ART Algorithm

Table 2. Result of Fuzzy ART algorithm (without Inference Rules)

Disease(Korean/Western)	Agreement
Kang-Hwa-Sang-Yom(Conjunctivitis)	2/4
Gan-Huh-Jung(Viral Hepatitis)	0/12
Pung-Han-Hyul-Tong(Colesystitis)	0/7
Dam-Huh-Jung(Hydrops of Galbaladder)	0/9
No-kwon-Sang(Neurastenia)	2/8

When fuzzy inference rules are applied, five most probable diseases are extracted as shown in Figure 14 and the number of agreed symptoms for those diseases is shown as Table 3. Names of diseases are represented by Korean medicine first and then corresponding western medicine disease name in parenthesis. This analysis is done by Korean traditional medical doctor. As one can see the comparison in Table 3, fuzzy inference rules play a critical role in our system to have a more accurate diagnosis.

Figure 14. Result Screen by Fuzzy ART and Fuzzy Inference Rules

Table 3. Result of Fuzzy ART algorithm (with Inference Rules)

Disease(Korean/Western)	Agreement
Kang-Wha-Sang-Yom(Conjunctivitis)	3/8
Sang-Cho-Wha	3/8
Migraine	3/11
Infectious	3/14
Yung-Hyul-Huh-Son (Multiple Mayeloma)	3/16

5. Conclusions

The proposed health pre-diagnosis system consists of two major parts - disease-symptom database and classification/learning algorithm. We construct the database based on KCD (Korean Standard Causes of Death Disease Classification Index) with Korean traditional medicine based on government reports and gathered information is verified by the KTM doctors for their validity.

This system gives the users instant access to information about the target diseases with causes and treatments from given user's symptoms. It will be available to the users in the shape of smartphone applications (app) as well as web applications. We partner with app evangelist to implement a successful shift to sustainable diagnosis system.

References

[1] A. G. Karegowda, A. S. Manjunath and M. A. Jayaram, “Application of genetic algorithm optimized neural network connection weights for medical diagnosis of Pima Indians diabetes,” International Journal on Soft Computing, vol. 2, no. 2, **(2011)**, pp. 15-23.
[2] M. Fathi-Torbaghan and D. Meyer, “MEDUSA: a fuzzy expert system for medical diagnosis of acute abdominal pain,” Methods of Information in Medicine, vol. 33, no. 5, **(1994)**, pp. 522-529.
[3] Y. S. Kim, Editor, “Dong-eui-bo-gam”, Solbit Publications, **(2003)**.
[4] C. G. Hong, “Complementary and alternative medicine in Korea: current status and future prospects,” The Journal of Alternative & Complementary Medicine, vol.7, no.1, (2001), pp.33-40.
[5] S. J. Lee, Editor, “Dong-eui-bo-gam by Symptoms,” Obi Enterprise, **(2004)**.
[6] C. H. Lee, Editor, “Chinese Medicine”, Minjoong Seogwan, **(1999)**.
[7] K. T. Han, H. C. Kim, J. Y. Ko and C. W. Lee, “Real-Time Intrusion Detection using Fuzzy Adaptive Resonance Theory,” The Journal of Korean Institute of Information Scientists and Engineers, vol. 28, no. 2, **(2001)**, pp. 640-642.
[8] K. B. Kim, M. N. Kim, Y. W. Woo, H. C. Noh and S. H. Shin, “Self Health Diagnosis System of Oriental Medicine Using Fuzzy Inference Method,” Proceeding of Winter Conference, Korea Society of Computer Information, vol. 18, no. 1, **(2010)**, pp.207-211.
[9] K. B. Kim, M. N. Kim, S. K. Cho and H. C. Noh, “Self Health Diagnosis System of Oriental Medicine Using Enhanced Fuzzy ART Algorithm,” Proceeding of Summer Conference, Korea Society of Computer Information, vol. 17, no. 1, **(2009)**, pp.329-332.
[10] K. B. Kim, S. Kim and K. B. Sim, “Nucleus Classification and Recognition of Uterine Cervical Pap-Smears Using Fuzzy ART Algorithm,” Lecture Notes in Computer Science, LNCS 4247, Springer, **(2006)**, pp. 560-567.
[11] K. B. Kim, Y. W. Woo and J. S. Kim, "Self Disease Diagnosis System Using Enhanced ART2 Algorithm,” The Journal of the Korea Institute of Maritime Information & Communication Sciences, vol. 11, no. 11, **(2007)**, pp.2150-2157.

Authors

Kwang Baek Kim, He received his M.S. and the Ph.D. degrees in Department of Computer Science from Pusan National University, Busan, Korea, in 1993 and 1999, respectively. From 1997 to present, he is a professor, Department of Computer Engineering, and Silla University in Korea. He is currently an associate editor for Journal of The Korea Society of Computer and Information, and The Open Artificial Intelligence Journal (USA). His research interests include fuzzy neural network and application, bioinformatics and image processing.

Doo Heon Song, He received a B.S. degree in Statistics & Computer Science from Seoul National University and M.S. degree Computer Science from the Korea Advanced Institute of Science and Technology in 1983. He received his Ph.D. Certificate in Computer Science from the University of California in 1994. Form 1983-1986, he was a researcher at the Korea Institute of Science and Technology. He has been a professor at the Department of Computer Games, Songdam College, Korea, since 1997. His research interests include ITS, machine learning, artificial intelligence, medical image processing, cognitive, and game intelligence.

