Advanced Science and Technology Letters Vol.46
(Games and Graphics 2014), pp.265-268
http://dx.doi.org/10.14257/astl.2014.46.56

Personal Identification System using BP Algorithm in Health Information Exchange System

Kyu Nam Choi[1] and Taeg Keun Whangbo[1],

[1] Department of Computer Science, Gachon University,
1342, Seongnamdaero, Sujeongnam-si, Gyeonggi-do 461-701, Korea,
robertchoimail@gmail.com, tkwhangbo@gachon.ac.kr

Abstract. It is hard to find out what is necessary among all the medical information of patients which is retained by various medical institutions within the health information exchange system. It is because each institution uses its own contents when retaining the basic information of their patients. In Korea, the identifier which distinguishes each patient was the resident registration number. However, as the concept of personal information protection applies to the medical industry, it is now considered undesirable to use the resident registration number and therefore it is required to have a new system which distinguishes patients without their resident registration numbers. This thesis is intended to design a system which primarily narrows the range of the patients based on their basic and medical information using their names and dates of birth and then secondly increases the accuracy of this classification using their gender, postal codes, contact numbers including mobile and home, types of insurance and medical areas to which the patients apply. Especially, BP Algorithm is used for similarity calculation by input values.

Keywords: BP Algorithm, Health Information, eMPI, Personal Identification, Health Information Exchange

1 Introduction

Recently, various systems for the health information exchange are being actively developed. Microsoft developed a system called HealthVault and Google also launched a service called Google Health. In Australia, NEHTA (National E-Health Transition Authority) was established as part of the federal government in 2005 and since then the authority carries out activities of establishing and distributing the integrated infrastructure and the relevant standards to deal with the compatibility issues among various healthcare systems and enhance the use of the patient- and supplier-oriented health information. According to the Hype Cycle for Healthcare Provider Application and System by Gartner in 2009, the health information exchange is now receiving its attention as one of the most important matters (1), and it is included in the nine development strategies for the next twenty years according to the report by the US National Library of Medicine in September, 2009 (2). It shows that

ISSN: 2287-1233 ASTL

the health information exchange is now drawing a great deal of attention from both industrial and academic fields.

The health information exchange can reduce the cost caused by duplicated preliminary examinations and prescriptions when patients change their hospitals or clinics and consequently reduce the time and medical cost. In addition, medical histories and allergic information can be obtained in advance to prevent the medical accidents and to provide customized medical services to patients.

A module to identify individuals needs to be in advance to the development of such health information exchange systems. A lot of studies have been carried out for peer-to-peer connection (3) and many types of algorithms were introduced and realized (4).

Such modules are distinguished as a Master Patient Index system called eMPI. As we still use the resident registration number, the relevant studies are not popular in Korea. However, as concerns for the use of resident registration numbers are growing due to the needs to protect personal information, it is now required to carry out studies on development of a health information exchange system without using the resident registration number.

2 Back propagation Algorithm

BP Algorithm is an algorithm which uses a delta rule generalized to a multi-layer perceptron as its learning rule. A hidden layer is located between the input and the output and each is connected to another with weight values while a bias value is located between the input and the hidden layer. It uses supervised learning as its learning technique. Multiplication and addition of the input and the weight values of the neuron is carried out several times and then the output can be achieved as a result of the input. As the output is not the expected result, weight values of the hidden layer are re-calculated in order to compensate the error values. Due to these characteristics, it is called back-propagation algorithm. In other words, the renewing direction of the weight values is from the output layer, the hidden layer in the middle and to the input layer. Figure 1 shows the typical architecture of BP Algorithm.

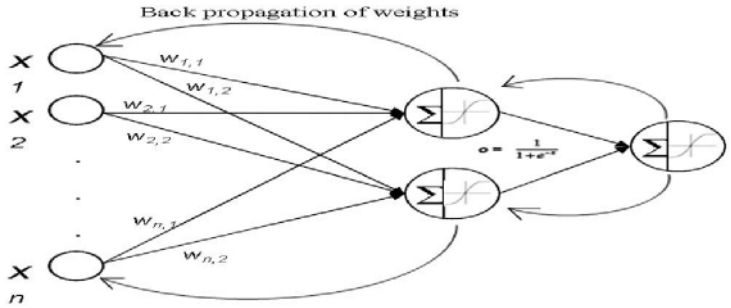
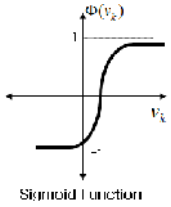


Fig. 1. Typical architecture of Back propagation Neural Network

At each stage, the sum of weight values are calculated using sigmoid function as described in the formula (1).



$$y_k = \Phi(v_k) = \frac{2}{1+\exp(-v_k)} - 1 \quad (1)$$

Reversely, the error ratio is calculated as described in the formula (2).

$$Eo = (To - Ao) \times sigmoidDerivative(Ao) \quad (2)$$

3 Proposed System

An algorithm proposed by this thesis to identify individuals applies BP Algorithm to primarily classify the patients with their names and dates of birth and then secondly improve the accuracy of the classification using their gender, postal codes, and contact information including mobile and home, types of insurance and medical areas to which they apply.

The output of BP Algorithm was classified into four classes and a table which defines the weight values of each class was used for learning. So the number of cases by the input value is classified into either one of 4 classes. Class 1 means the highest priority. The highest priority is given to situations where the type of insurance and the medical area are not matching. That is to say that the actual matching has no problem although those values are not matching. In other countries, they do not use the same patient information as ours in MPI algorithm and therefore, the probability for not having is high. Class 2 is configured where the home number is not matching with the mobile phone number. Mobile phone numbers and home numbers relatively have high missing rates. Class 3 is configured where the postal code is not correct and Class 4 where the gender is not correct. Once the environment variables for BP Algorithm are completely configured, configure the initial weight values. Those values are repeatedly calculated to reduce the error rate.

We measured the accuracy rate for the comparative evaluation between the proposed and existing algorithms. The testing result is shown in Figure 2.

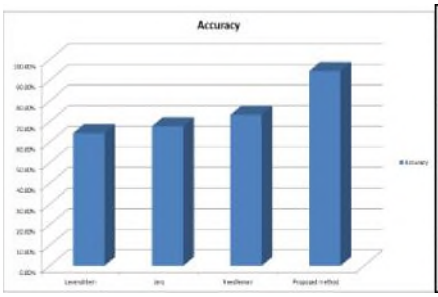


Fig. 2. Accuracy Test Result

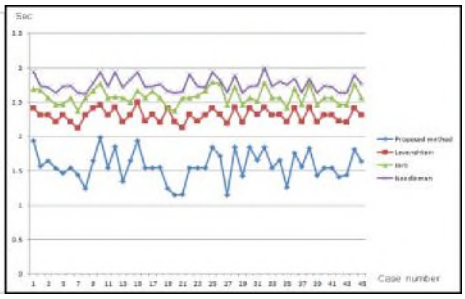


Fig. 3. Speed Test Result

Figure 3 illustrates the result of the speed testing for comparison with existing methods. The overall result indicates that the proposed method takes less time than the existing methods.

4 Conclusions

The proposed system enables the accurate patient matching using their names, dates of birth, basic information and patient information without their resident registration numbers. In addition, we can improve the accuracy based on the preliminary learning using a neuron network. However, we found out that entering inaccurate patient names resulted in no matching as the patients were primarily clustered with their names and dates. We will improve the algorithm based on studies on a string matching algorithm which is appropriate to two-byte characters of Asian regions in the following study.

Acknowledgement. "This work was supported by the Industrial Strategic Technology Development Program (10044791) funded by the Ministry of Trade, Industry and Energy (MOTIE) of Korea".

References

1. Gartner's Hype Cycle for Healthcare Provider Applications and Systems. (2009)
2. NLM. Charting a Course for the 21st Century: NLM's Long Range Plan 2006-2016. http://www.nlm.nih.gov/pubs/plan/lrpdocs.html Accessed August 24. (2009)
3. Fellegi I., Sunter, A.: A Theory for Record Linkage," Journal of the American Statistical Association, 1183-1210, December. (1969).
4. AnHai D.: Principles of Data Integration, pp. 95-117. (2012).

