International Journal of Multimedia and Ubiquitous Engineering
Vol.10, No.11 (2015), pp.293-302
http://dx.doi.org/10.14257/ijmue.2015.10.11.28

Analysis Based on Mining Improved BP Neural Network in Medical Quality Monitoring

Dina Xia

China Electronics Standardization Institute, Beijing, 100007
bjxiadina@sina.com

Abstract

Aiming at the deficient methods to predict the increasing medical malpractice, improved BP neural network is adopted to analyze and predict medical malpractice. The introduction of genetic algorithm into BP neutral network well overcomes the random initial weight value of BP neural network, network oscillations brought by determining network and that the network is easy to fall into local solution. In addition, it effectively improves generalization ability of neural networks. Simulation experiments show that improved BP neural network algorithm has higher accuracy in predicting medical malpractice in quality supervision.

Keywords*: BP neural network; genetic algorithm; monitoring over medical quality*

1. Introduction

Over the past decade, medical engineering with the development of advances in computer technology have made great progress, especially in the wake of medical information system in hospitals gradually put into use, patients see a doctor during the whole process using computer for storage, from data in a Data Set using a variety of data mining technology on various cases mutual relations and regular, and sums up the medical case [1]. Scholars study the data in the medical literature [2] proposed for large data storage, calculation, analysis, dealing with new issues, especially on data mining. Background of the articles analyze current data, describes the basic features of data and applications, combined with the medical field, discusses medical purpose and main methods of data analysis. Literature [3] presented medical data analysis and predictive modeling, improved medical data analysis systems and hospital data transfer protocols between database, improve overall system of medical data transmission efficiency and forecast accuracy. Literature [4] medical data analysis service is designed to provide health care cloud platform and deployment. Secondly, the application provides a platform for multiple tenant requests scheduling policy based on Queuing models is presented. Experimental results verify that the frame design efficiency of the method. Literature [5] combining cloud computing a Hadoop-based ecological environmental building of medical data mining platform architecture, described the function of each layer, including the base layer, platform, functionality and the business tier, with a view to provide data analysis and mining of medical industry with new ideas. Literature [6] established a pretreatment method, at the time of medical data mining, the dependent variable (skew the distribution of a continuous variable) conversion as categorical variables, so as to be more scientific and reasonable results. Literature [7] presents a weighted, based on risk and prevention of medical data mining model algorithms, enabling them to make better mining effects.

On the basis of the above literatures, this paper adopts improved BP neural network to

ISSN: 1975-0080 IJMUE

analyze and predict medical malpractice. The introduction of genetic algorithm into BP neutral network well overcomes the random initial weight value of BP neural network, network oscillations brought by determining network and that the network is easy to fall into local solution. In addition, it effectively improves generalization ability of neural networks. Simulation experiments show that improved BP neural network algorithm has higher accuracy in predicting medical malpractice in quality supervision.

2. Overview of Medical Data Mining

2.1 Definition of Requirements

During this stage, analyze demands and find resources, define demands, determine medical targets and medical groups to get to know solutions to solve this demand. Evaluate current resources and then determine target of mining, formulate mining plans and extract a certain amount of sub-database from original data to establish data mining base.

2.2 Pre-Treatment of Data

Determine data sources of mining, including data source in database or external data of database that are gathered independently, and list types of data, including property, size, and format so as to describe data source and realize the purpose of answering data mining with the initially obtained data. It can also confirm the initial hypothesis or explore new characteristics. Then, choose data sources in two ways: one is to collect samples of data space, and at this time, choice of data is random; another is to collect samples of featured space, and data with some specific features can be chosen.

2.3 Data Mining

Use appropriate data mining algorithms for data analysis, which is a critical medical knowledge discovery process step. The method of data mining includes concept description, association analysis, classification and prediction, clustering analysis, trend analysis, outlier analysis and error analysis. Mining results in practical application requires repeated verification, to test its rationality, between the need to establish a broad exchange of experts and the various experimental results and their medical experience for confirmation to verify the reasonableness of mining methods.

2.4 Analysis and Evaluation

Assess whether the knowledge is accurate and compare them with the initial research target. The conclusion must be understandable for users. In the medical field also need to assess whether knowledge out innovative, useful, conclusion can be explained on the medical, and consistent with the original goals of data mining.

2.5 Determine Implementation Scheme

After the model is established and verified, implement the model with two methods: one is to provide them to analysts for reference and proposes corresponding implementation scheme after observing and analyzing this model; another is to apply this model to different databases.

3. Mining Algorithm Based on Improved BP Network

3.1. Proposition of Problems

Due to genetic algorithm of search not rely on gradient information, also not need solution function can micro-, only need solution adapted degrees function in constraints conditions Xia can solutions. and genetic algorithm has global search of characteristics, with genetic algorithm optimization neural network of connection right and network structure, can better to overcome BP neural network of initial right value of randomness and network structure determine process in the by brings of network oscillation, and network very easy into local solutions problem, and effective improve neural network of generalization capacity. Therefore, using genetic algorithm for global search feature, look for the most suitable network connection power and network structure to change BP algorithm of gradient-dependent instructions, so as to achieve the network structure and network optimal allocation of the initial connection weights and thresholds. In the switching coefficient is introduced in this article, k is used to represent the neural connections between network nodes evaluate to 1 or 0, to say there is any connection between the two nodes, respectively. Mathematical description of genetic-neural network for optimization problems are as follows:

$$E_1 = \sum_{k=1}^{N_1} \sum_{t=1}^{n} [y_k(t) - \hat{y}_k(t)] \tag{1}$$

Herein, E_1 is the total error of network, $y_k(t)$ is theoretical output of network and $\hat{y}_k(t)$ is the actual output of network. Herein, $\hat{y}_k(t)$ is as follows:

$$\hat{y}_k(t) = \sum_{i=1}^{m} w_{ij} \square x_i(t) + \theta_j \tag{2}$$

Suppose $E_2 = \dfrac{\sum_{k=1}^{N_1} \sum_{t=1}^{n} [y_k(t) - \hat{y}_k(t)]}{N - N_1}$ is average MSE of test samples, showing estimation of reliability of network's output data. The smaller the value of E_2, the more reliable the network output is. Otherwise, network's output is not so reliable.

During the process of applying neural network, as the properties of approximation samples cannot be clearly know, so even if the network error E_1 is zero, it cannot guarantee that E_2 can reach the requirements, and the case is often that E_1 is very small, but E_2 cannot meet the requirement, thus the network finally loses its value, so in the design of the algorithm, E_2 should be smaller than a given error ε, so as to make network error reliable.

3.2. Design of Algorithm

The algorithm in this paper through using genetic algorithm in by BP network preliminary determine of basic solutions space Shang (network connection right and neurons valve value of take value range), through on gene of select, cross variation operation. on sample individual constantly preferred evolution, until after evolution (total of evolution algebra by initial Shi given) generation, selected individual in the adapted degrees maximum of individual to determine network of structure and network of initial

right value and valve value.

(1) Gene Coding

Code string encoded in this article by the switching coefficient coding, weighting coefficient coding consists of two parts. Switch coding hidden layer neurons and neurons in input and output connection status, switch coding string length l_1 can be determined implicitly by the initial number of neurons. Actually, the number of hidden layer neurons is determined by switching the number of coefficients in the 1. Weighting coefficients represent network connections right, float-encoding, string length $l_2 = (m+n) \times l_1$, where m is the number of input nodes and n is the number of output nodes). Coding according to a certain order of cascading into a long string, each string corresponds to a set of network structure and connection weights.

(2) Determine Space of Basic Solution

Three-layer BP network is used to preliminarily determine basic solution space (ranging of network connection right and neuron threshold). Firstly, set the times of network training and the network training error ε_1, and input training samples for training. Afterwards, input test samples to get the error ε_2. When both the ε_1 error ε_2 and are satisfactory, regard the maximum and minimum value of connection right value as $u_{\max}$ and. $u_{\min}$ with [$u_{\min}$, $u_{\max}$]as the basic solution space of connection right.

(3) Calculate Fitness Function

In genetic algorithms, the fitness function is only required for input can be calculated to compare the non-negative result. Therefore, this network error function is used as a fitness function, and that error the individual adaptation of small, concrete expressed as

$$F(x) = \sqrt{\sum_{k=1}^{N_1} \sum_{t=1}^{n} [y_k(t) - \hat{y}_k(t)]^2} \tag{3}$$

Calculate the fitness of each individual in the group, by switching coefficients are the number of hidden nodes of the network, by the weight of the digital network connections right, enter the training samples in accordance with (3) calculate the fitness of each individual.

(4) Cross-over and Mutation Factor

Cross is by larger of probability from groups in the, random select 2 a individual, Exchange this 2 a individual of some bit, exchange of purpose is produced new of gene combination, to limit genetic material of lost. Variation is to smaller of probability on groups in some individual of bit for change, in real coding is on some individual of bit in (0, 9) produced a random number instead of original individual of bit. Variation of purpose is protection some adapted degrees low of individual in excellent gene, prevent found excellent process in the early convergence Yu not mature district.

Switching coefficient encoding crossover and mutation using classical genetic algorithm in a single point crossover, uniform mutation of Method, in the crossover and mutation operation, when a neuron deleted, the corresponding weighting factor related to coding and coding threshold coefficient is set to 0, which means that the network connection is dropped; and operation of a neuron increases, then the random initialization related weighting factor coding and neural coding threshold due weight coefficient coding and neural encoding using floating-point threshold coding, we need to design a new crossover and mutation operator.

Cross individuals after selection with P_c as the probability. Suppose the cross is made between the individual i and $i+1$, and the cross factor is as follows:

$$\begin{cases} x_i^{t+1} = c_i \square x_i^t + (1 - c_i) \square x_{i+1}^t \\ x_{i+1}^{t+1} = (1 - c_i) \square x_i^t + c_i \square x_{i+1}^t \end{cases} \tag{4}$$

Herein, x_i^t and x_{i+1}^t is a pair of individuals before crossing, x_i^{t+1} and x_{i+1}^{t+1} is a pair of individuals after crossing, and c_i is a random number between 0 and 1. Make mutation towards individuals after crossing with p_m as the probability, suppose the mutation is made toward individual i, and then the mutation factor is as follows:

$$x_i^{t+1} = x_i^t + c_i \tag{5}$$

x_i^t is the individual before mutation, x_i^{t+1} is the individual after mutation and c_i is a random number evenly distributed serving to guaranteeing that individuals after mutation are still within the range of search.

For crossover operation of weighting coefficient coding and neuronal threshold coding, multi-point crossover genetic operation is adopted. This can further increase the dispersibility of genetic search, thus making it converge into precision needed more rapidly. Multi-point crossover cross-over points randomly generated according to a given probability, intersection position is randomly generated. Weighting coefficients for coding and neuronal coding mutation threshold multi-point adaptive mutation are to make fitness a large individual variation in the smaller range, leaving the adaptation of small individual variation in the larger context, the introduction of mutation concept of temperature, this concept is similar to the concept of simulated annealing algorithm temperature variation solution temperature is defined as follows:

$$t = \frac{f_{\max} - f(s)}{f_{\max}} \tag{6}$$

Herein, f(s) refers to fitness of the individual s, $f_{\max}$ is the maximum fitness of problems to be solved, which only requires a rough upper limit, or the maximum fitness of current group can be taken as $f_{\max}$; after the concept of variant temperature is introduced, such variation method can be used to select an individual v_i from each individual v for positive distribution of individual v_i' after variation.

4. Analysis Data Mining in Medical Quality & Security Monitoring

4.1. Monitoring and Analysis of Medical Data

Taking construction Fuzhou Public Health Emergency Network system as the opportunity, Fuzhou Bureau of Health fully refers to advanced foreign medical health monitoring model and their mature management experience to set up medical quality & security monitoring system covering all the medical institutions above Grade A Level Ⅱas well as Health Monitoring Center of Bureau of Health above the county level, which is a bold exploration and attempt towards Fuzhou medical security monitoring model. Medical quality security monitoring system is made up of five sub-systems: information collection, information analysis, information precaution, information feedback and information assessment with collecting information about malpractice of medical security as the basis, analysis as the core, precaution as the means and promoting patients' health as the target. Its model of operation is as shown in Figure 1:

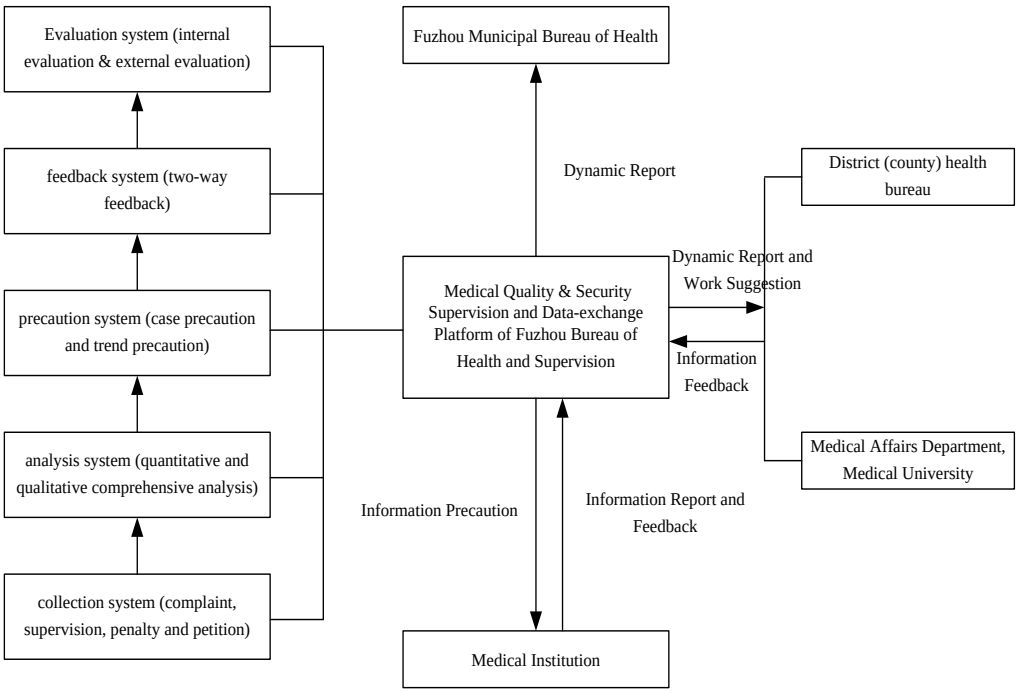

Figure 1. Model of Operation

4.2. Classification of Doctor-Patient Relationship

During the process of extracting and transforming data of medical disputes, as well as cleaning and loading, the following characteristic properties of data can be analyzed.

(1) Analyze patients' characteristics. Characteristic properties of patients include their age, gender, occupations, residence, previous illness, type of illness as well as their consequences, type of medical insurance, medical costs and past medical disputes, *etc.*

(2) Analyze characteristics of medical staff. Characteristics properties of medical staff include age, gender, title, department, employment age of doctors and nurses, as well as previous medical disputes, *etc.*

(3) Analyze characteristics of medical institutions. Characteristic properties of medical institutions include nature and grade of medical institutions, relevant rules and regulations, history of similar medical disputes, and history of medical malpractice, *etc.*

(4) Analyze characteristics of disposing medical disputes. Characteristic properties of medical disputes include contents of medical disputes, complaint channels, solutions, quality of medical disputes, required compensations, actual compensations and relevant media reports, *etc.*

4.3. Decision Support of Improving Medical Quality

Classify medical malpractices and adopt corresponding $x_1, x_2, \cdots x_k$ as the training set to establish neural network model according to different types, and use $x_{k+1}, x_{k+2}, \cdots x_{k+s}$ as the test sample to verify the correctness of neural network modeling, which is as shown in Table 1. According materials, this paper simulates and chooses medical malpractice data of Fuzhou in recent ten years and then obtains results as in Table 2.

Table 1. Learning Samples and Test Samples

Sample Type	Sample No.	Output			Input
		x_i	x_{i+1}	x_{i+2}	x_{i+3}
Learning Samples	0	0.7504	0.7750	0.8132	0.8630
	1	0.7750	0.8132	0.8630	0.9232
	2	0.8132	0.8630	0.9232	0.9120
	3	0.8630	0.9232	0.9120	0.8828
	4	0.9232	0.9120	0.8828	0.9162
	5	0.9120	0.8828	0.9162	0.9999
	6	0.8828	0.9162	0.9999	0.9276
	7	0.9162	0.9999	0.9276	0.7803
	8	0.9999	0.9276	0.7803	0.7747
	9	0.9276	0.7803	0.7747	0.8124
	10	0.7803	0.7747	0.8124	0.8600
	11	0.7747	0.8124	0.8600	0.8882
	12	0.8124	0.8600	0.8882	0.8395
Test Samples	13	0.8600	0.8882	0.8395	0.7697
	14	0.8882	0.8395	0.7697	0.7558
	15	0.8395	0.7697	0.7558	0.7648

Table 2. Comparison of Predicted Results between Standard BP Neural Network and Improved BP Neural Network

Year	Simulated Medical Malpractice (people)	Predicted Value of Standard BP Neural Network (People)	Difference (People)	Relative Error (%)	Predicted Value of Improved BP Neural Network (People)	Difference (People)	Relative Error (%)
2004	9460	8857	-603	-6.374	9247	-213	-2.252
2005	5650	7132	1482	26.23	6032	382	6.7611
2006	9100	3984	-5116	-56.22	5782	-3318	-36.462
2007	3600	5552	1952	54.2222	4129	529	14.6944
2008	4139	1980	-2159	-52.162	3682	-457	-11.041
2009	6000	5579	-421	-7.017	5925	-75	-1.25
2010	8000	8233	233	2.9125	7924	-76	-0.95
2011	8900	2764	-6136	-68.944	3781	-5119	-57.517
2012	2080	3280	1200	57.6923	1491	-589	-28.317
2013	7900	9187	1287	16.2911	8123	223	2.8228
2014	5800	4064	-1736	-29.931	5219	-581	-10.017

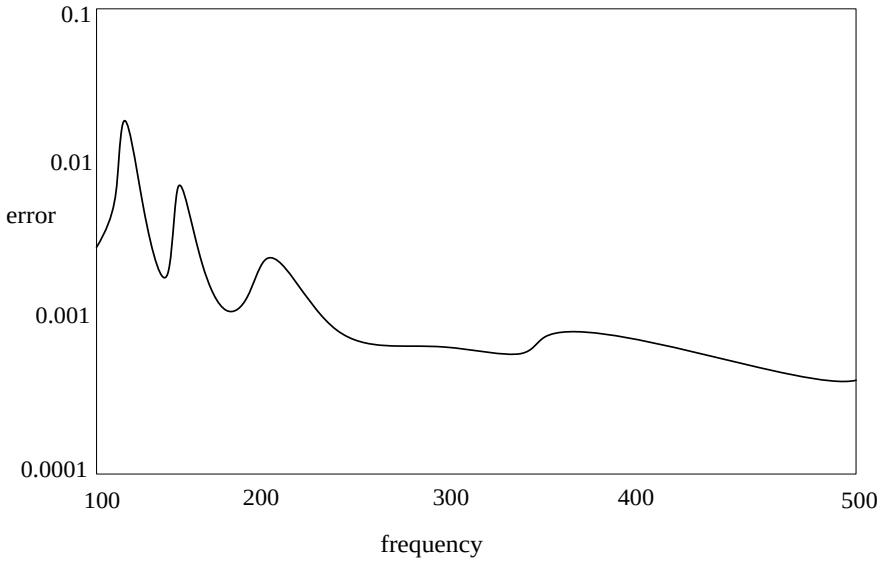

Figure 2. Error Curve of Improved BP Neural Network

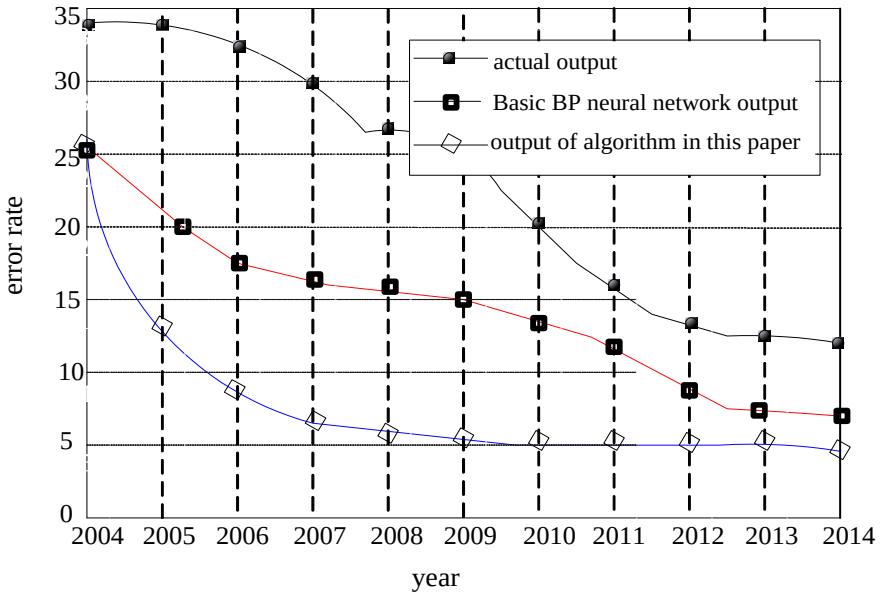

Figure 3. Comparison of Simulation Results

It can be seen from simulation results of Table 1-2 that the maximum relative error of standard BP neural network is 68.944% while that of improved BP neural network is 57.517%, indicating output results of improved BP neural network are more accurate with fewer errors. Thus it can effectively predict the occurrence of medical accidents. Error curve of improved BP neural network is as shown in Figure 2, and it can be seen that after 500 times of training, stable value of the curve is reached, indicating that expected indexes are reached. Figure 3 shows comparison of simulation result, and in the predicted results, the maximum relative error is in the year 2006, difference between actual medical malpractice and standard BP neural network's output value is 17% while that between improved BP neural network output and actual passenger capacity is 24%. This shows that improved BP neural network has more accurate prediction results with higher accuracy.

5. Conclusion

This paper adopts improved BP neural network to analyze and predict medical malpractice. The introduction of genetic algorithm into BP neutral network well overcomes the random initial weight value of BP neural network, network oscillations brought by determining network and that the network is easy to fall into local solution. In addition, it effectively improves generalization ability of neural networks. Simulation experiments show that improved BP neural network algorithm has higher accuracy in predicting medical malpractice in quality supervision.

References

[1] H. Hong and L. L. Juan, "The Characteristics, Process and Methods of Medical Data Mining", Value Engineering, vol. 32, no. 1, **(2011)**, pp. 166-167.
[2] Z. B. Ji, "Big data analysis and its application in medical field", Computer Education, vol. 10, no. 7, **(2014)**, pp. 24-28.
[3] T. Y. Chi and H. Liang, "A medical data analysis model based on SVM", Journal of Northeast Normal University(Natural Science Edition), vol. 47, no. 1, **(2015)**, pp. 77-82.
[4] H. X. Huan, C. Mei and L. Hui, "Data Analysis As A Cloud Service Based on Healthcare Cloud", Computer & Digital Engineering, vol. 43, no. 4, **(2015)**, pp. 656-659.
[5] G. H. Song 等, "Medical data mining platform based on Cloud Computing", Journal of Medical Informatics, vol. 34, no. 5, **(2013)**, pp. 7-12.
[6] S. Pei and Z. J. Kai, "Application of clustering analysis in data mining of medical expenses", South China Journal of Preventive Medicine, vol. 38, no. 1, **(2012)**, pp. 18-23.
[7] Z. J. Peng, H. J. Feng and M. Lei, "Medical Data Mining Algorithm Based on Optimal Risk and Prevention Model", Computer Engineering, vol. 37, no. 22, **(2011)**, pp. 32-36.

Author

Dina Xia (1979.03-), female, Master, Her research direction is electronic information technology, the Internet of things, computer technology and data mining.