Genetic Neural Network for Traditional Chinese Medicine

Yun Wu
The Graduate School of Artificial Intelligence, XiaMen University
FuJian XiaMen
361005, China
E-mail: yininwu@126.com

Changle Zhou
The Graduate School of Artificial Intelligence, XiaMen University
FuJian XiaMen
361005, China
E-mail: dozero@xmu.edu.cn

Zhifeng Zhang
Shanghai University of Traditional Chinese Medicine
ShangHai
201203, China
E-mail: rchbt@163.com

Abstract—**The research into the impersonality and information of the Traditional Chinese-medicine Diagnosis (TCMD) is a key work to realize the Traditional Chinese-medicine (TCM) medicine examination system. So the paper studies the application of Artificial Neural Networks (NN) to TCM and uses the Genetic Algorithm (GA) to optimize the ANN, which is used to the diagnosis of TCM. The initial structure of the ANN is arbitrary and the GA is able to find a near optimal connection weight and structure of the ANN for the TCMD. In this paper, we emphatically describe t the GA to optimal the connection weight of the NN — Genetic NN, which has a relatively fixed three-layer and was used to the TCM eight principal syndromes by the tongue feature.**

Keywords—— TCM (Traditional Chinese-medical); tongue feature; TCM eight principal syndromes; NN(Neural network)、Genetic Algorithm.

I INTRODUCTION

Tongue diagnosis is a very important content in the Sizhen of TCM —Inspection, and is a characteristic diagnostic method. The clinic study has proved, during the developing of the diseases, the change of the tongue is rapid and brilliant. It is like a mirror of the bowels, and the pathological changes of the bowels can reflect from the change of the tongue features. So it becomes the important evidence for the diagnosis. But because of the lack of the impersonality and information of the TCMD, it becomes very difficult on the road of the TCM modernization, and all the time, it is a blankness in the field of the medicine examination. But comparing others diagnostic methods, it has got some research result on the impersonality and information of the tongue diagnosis, which provide the impersonal information criterion for the research and development of the TCM diagnostic methods based on the tongue features and promote the course of the TCM in the world. It also can develop a new field for the TCM medicine examination system.

The TCM medical examination systems integrate these main technologies, such as logic, NN, GA, and others soft computing. Its main theory is based on these: First the characteristics of the tongue feature, and get the logical rule. Second construct the NN for every the TCM diagnosis rule. Because the common ANN has very weak auto adaptability, which can not satisfy the variety of the TCM clinic, So At last use the improved GA to optimize the NN(namely GNN), and form the TCMD NN database. The TCMD NN database is information source of the TCM medical examination system and the kernel technology. But the improved GA is the key technology to form the TCMD NN database.

The GA the paper put forward that used to optimal the TCMD ANN is an improved traditional GA. At first, to achieve a good offspring, know data are divided into two sets namely training set and validation set, both of which are used to compute the fitness of the NN. Then compute the fitness of each individual, namely one NN, and according to the fitness values, rank the NN from the best to the worst, and memorize the best one and the worst one. According the rank of the NN, put the former n individual directly copied to next generation, the remaining offsprings come from the crossover operation. Void the local excellent result via a crossover of the best individual and the worst individual. In the mutation, we use the PSO[4], namely use the fitness value of the best individual and the worst individual via the mutation function to operate the mutation, which is in favor of the convergent. And in order to research the best solution of the whole solution space, the multi-GA is considered, namely get the best solution via select the best fitness from several GA computing.

This paper is organized as follows. In section II, we present three-layer NN, which base on the characteristic of tongue diagnosis and the data format of TCM eight principal syndromes, and introduce the improved the general sketch our GA design, which combine the thought of real coding, arithmetic recombination, PSO, etc. In section II, we test the improved GA on tongue figure-based TCM eight principal syndromes data, which form is presented in section II, and describe the simulation results. Conclusions and future work has been disused in section III.

0-7803-9422-4/05/$20.00 ©2005 IEEE

1) The ANN based on the TCM tongue features

(1) The structure of the tongue figure-based TCM eight principal syndromes ANN

For using the improved GA to optimize the value and the structure of the tongue figure-based eight principal syndromes NN, we initialize an arbitrary ANN, which has three layers. The three layers are the input layer, the hidden layer and the output layer. The input layer has the value contact with the hidden layer, and the hidden layer has the contact with the output layer. The construct shows in figure1.

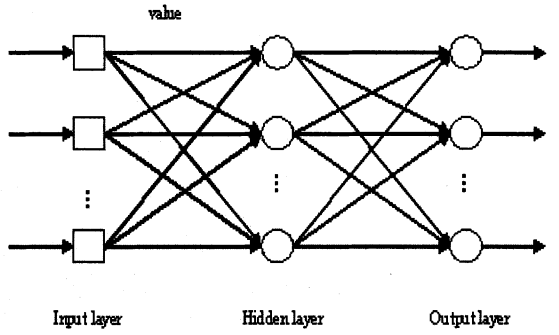

Figure 1. the three-layer feed-back ANN's structure

The number of the input of the training set confirms the number of the input cell. The number of the hidden nerve cell is confirmed by our experience. The number of the result confirms the number of the output cell.

(2) The data structure of the tongue feature-based TCM eight principal syndromes

we use the tongue figure(or tongue symptoms) to be the example to diagnosis the "Eight principal syndromes", which include eight syndrome, such as "Exterior、Interior、Cold、Hot、Deficiency、Excess". The detail context can been seen this table.

syndrome	value	Tongue figure		
		Body		
		color	shape	state
exterior	1	Light red	constancy	constancy
interior	2		inconstancy	inconstancy
cold	3	light white		
hot	4	crimson		
deficiency	5	light white、crimson	tender	
excess	6	blueness purple	old	
default	7			

Table 1 (a)

syndrome	value	Tongue figure	
		Coating	
		color	proper
exterior	1	white、yellow	thin
interior	2		thick
cold	3	white	moist
hot	4	yellow	dry
deficiency	5		exfoliative
excess	6		thick、greasy
default	7		

Table 1 (b)

Explain to Table 1 1. the blank in the table means the syndrome has not distinct symptoms or the symptoms can not help much to confirm the syndrome. 2. the default means the symptoms is not enough, so the syndrome can not be ensured. 3. this table is made based on the reference book [1] in the catalog, in order to confirm that the technology of the ENN is valid.

(3) Date the symptoms, such as

Tongue Color: light red{1}; light white{2}; crimson{3}; blueness purple{4}

Tongue Shape: constancy{1}, inconstancy{2}, tender{3), old{4}

Tongue State: constancy{1}, inconstancy{2}

Coating Color: white{1}, yellow{2}

Coating Proper: thin{1}, thick{2}, moist{3}, dry{4}、greasy{5}, exfoliative{6}

If there are not symptoms to the syndromes, the value that the symptom is corresponding in the table is zero.

(4) Construct data set of the input stylebook and output stylebook.

[Tongue color, Tongue shape, Tongue state, Coating color, Coating Proper] /[Eight principal syndromes], according the table 1, the stylebook data like this:

Exterior: [1, 1, 1, 1, 1]/[1]; [1, 1, 1, 2, 1]/[1]

Interior: [0, 2, 2, 0, 2]/[2]

Cold: [2, 0, 0, 1, 3]/[3]

Hot: [3, 0, 0, 2, 4]/[4]

Deficiency: [2, 3, 0, 0, 6]/[5]; [3, 3, 0, 0, 6]/[5]

Excess: [4, 4, 0, 0, 2]/[6]; [4, 4, 0, 0, 5]/[6]

default: [0, 0, 0, 0, 0]/[7]

In order to increase the quantity of the input training set, we pick up some data form the inference book[1] and the clinic data of the author[3], which has six kinds of TCM eight principal syndromes' data.

2) The GA for optimizing the TCMD ANN

The main steps of the GA for optimizing the TCMD ANN are:

(1) Real-valued two-matrix encoding. A network is represented as a real-valued matrix, in which c_{ij} is zero or is a very small number if there no link from the input neuron i to the hidden neuron j, which can encode both the structure

the weights.

(2) Initialization. Using the conventional method, randomly generate an initial population of networks with number of input neurons and the hidden neurons, the connection weights assigned at random. Before into next step, each network is partially trained, this can avoid overstraining.

(3) Fitness function. Known stylebook set is divided into two sets namely training set (Ts) and validation set (Vs). After partially trained, compute the fitness of each NN, and rank them from the best to the worst memorized the best NN of this generation and modify the NN of the all generations. The fitness function is:

$$f(x) = \alpha \times E(Ts) + (1-\alpha) \times E(Vs)$$

$$E(T) = \frac{1}{MSE\ (x)}$$

(4) Genetic operation. Directly copy K individuals (K usually is twenty percent of the population of the NN) have best fitness to next generation, for keeping the best gene. The remaining offspring comes from the crossover operation. For escaping the local excellent result, bringing one offspring by crossover between the best individual and the worst individual.

① Selection. According the sorted individuals number, the ith individual is selected with a probability of

$$p_i = 1 - \frac{\sqrt[4]{i}}{\sum_{i=1}^{m} \sqrt[4]{i}}$$

the$\alpha$is a random number, and if theα<p_i, the individual is selected.

② Crossover. Select two individual, and crossover the two individual via the mathematic recombination. Assuming that $X = \{x_1,\ x_2,\ldots,\ x_n\}$ and $Y = \{y_1, y_2,\ldots,\ y_n\}$ are two individual for crossover, and the x_i, y_i are the genes in the two individual. The crossover is implemented as follows: $z_i = h_i \times x_i + (1-h_i) \times y_i$

the $h_i \in (0,1)$ is a random number。 Subsequently, each NN is partially trained.

③ Mutation. Select individuals and conduct mutations according to the current mutation rate p_m. Considering the method of the TCM case learning which are always applied in the TCM expert systems, so the mutation uses the thought of PSO, namely according the fitness of the best individual and the worst individual, randomly select two individuals, such as i and j, then mutation is implemented as follows:

$$x'_{i,j} = x_{i,j} + \phi_1 \times (f(p_g(t)) - f(x_{i,j}))$$
$$+ \phi_2 \times (f(p_d(t)) - f(x_{i,j}))$$

the $p_g(t)$ is the best evolutionary NN of the all generation, $p_d(t)$ is the best evolutionary NN of this generation, $\phi_1, \phi_2 \in [-1,1]$ are random number. This method simulates the thought of TCM case learning, and is propitious to the convergence of the GA.

(5) Ending condition. In the GA computing, if neither the maximal evolutionary epoch is reached nor the unchanged times of the fitness of the best NN is reached, go to (2). If neither the performance of the best NN meets the requirement or the unchanged times of the fitness of the best NN, which get from one GA computing is reached, go to (1).

II. MAIN RESULTS

According to the data form of the tongue figure-based eight principal syndromes, we initial a three-layer feed-back NN, the output layer has five input cells, the hidden layer has ten nerve cells at initially, and the output layer has seven output cells.

When the GA is computing, the partial training graphs of the TCM eight principal syndromes NN is different, such as:

1) the prophase of the GA:

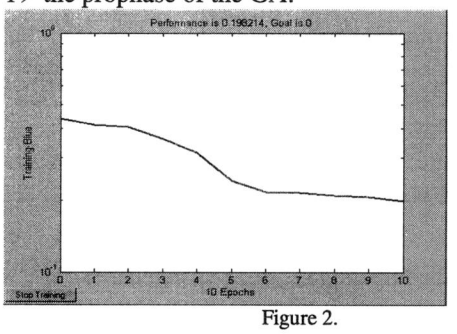

Figure 2.

2) the metaphase of the GA:

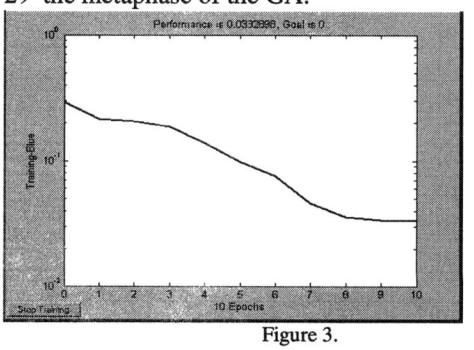

Figure 3.

3）the anaphase of the GA:

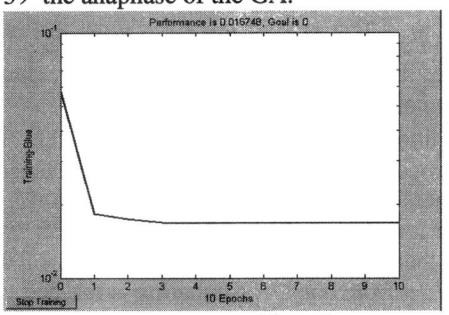

Figure 4.

The partial computing result of the optimized the tongue figure-based TCM eight principal syndromes NN is following:

1）the partial computing result of training set:

Input data	Computing result	Exact result
[1,1,1,1,1]	1	1
[0,2,2,0,2]	2	2
[2,0,0,1,3]	3	3
[3,0,0,2,4]	4	4
[3,3,0,0,6]	5	5
[4,4,0,0,2]	6	6
[0,0,0,0,0]	7	7
[0,2,2,0,0]	2	2
[4,0,0,0,2]	6	6
[0,0,0,0,6]	5	5

Table 2.

2）the partial computing result of valuable set:

Input data	Computing result	Exact result
[3,0,0,0,4]	4	4
[4,4,1,1,5]	6	6
[3,1,1,2,4]	4	4
[2,1,1,1,3]	5	3
[1,0,0,0,0]	6	7
[3,0,0,2,0]	6	4
[4,4,1,0,0]	6	6
[0,3,0,0,6]	5	5
[0,1,0,1,1]	3	1
[1,2,2,0,0]	2	2

Table 3.

After optimized by the GA, the exact rate of the TCM eight principal syndromes NN is 100% to the training set, the exact rate to the valuable set is 60%, which is more excellent and more stable than the common ANN that is used to TCMD.

III. CONCLUSIONS

This paper discusses the GA optimize the ANN, which is used to TCMD—— TCM eight principal syndromes based on the tongue feature. Using the arithmetic recombination and the theory of the PSO to help the GA escape from poor local excellent value and improve the convergence. At the same time, in order to improve the computing correctness of the validation set, we make the computing value of the validation set act as one of factors, which are used to compute the fitness. After the simulating, the Genetic NN to TCMD has better performance than initial NN, specially in the adaptability and the stabilization. If the Genetic NN has automatic structure based on the stylebook data, it can perform better, which is our next work.

REFERENCES

[1] Wenfeng Zhu，Traditional Chinese Medicine Diagnostics [M]Beijing：Chinese Medicaments Press，2004

[2] Nenggan Zhang，TCM-SIRD:An Intergrated Aided System for Traditional Chinese Medicine Sizheng[J]，2004 IEEE International Conference on System, Man and Cybernetics，2004，（3864-3868）

[3] Xuezhong Zhou，TCMMDB:A Distributed MultiDatabase Query System and It's Key Technique Implement，2004，IEEE(1095-1098)

[4] Chia-Feng Juang，On The Hybrid of Genetic Algorithm and Particle Swarm Optimization For Evolving Recurrent Neural Network，2004，IEEE（2285-2288）

[5] Z.-H Tan，Hybrid evolutionary approach for designing neural networks for classification，ELECTRONICS LETTERS 22nd July 2004 Vol. 40 No. 15

[6] Chonghong Dong，Matlab, Neural Networksand Application [M]Beijing：National Defence Industry Press，2005