A Method of Remote Fault Diagnosis Based on Analytical Hierarchy Process

WANG Xiao-bin CHEN Wen-yu Sun Shi-xin
School of Computer Science and Engineering
University of Electronic Science & Technology of China
Chengdu Sichuan , China
cwy@uestc.edu.cn

LIU Jing-bo
School of Mathematics and Computer Science
Chongqing Three Gorges University
Chongqing , China
allahljb@163.com

***Abstract*—This paper presents a method for Neural Network Ensemble .In the method, five subsystems of classifier used four kinds of neural networks, such as SOM,PNN,LVQ,RBF. Those neural networks compute parallel which have been trained solely .The recognition result of subsystems, the expectation and variance between input pattern and goal pattern have been integrated by Analytical Hierarchy Process .In experiments, the proposed methods have been successfully evaluated using thirteen different datasets, it is more effective than the relative majority voting scheme . The integration method consumes little computing resource and the result of calculation accords with the actual conditions, which indicates that AHP is an efficient ensemble method.**

***Keywords*—Analytical Hierarchy Process ;neural network ; ensemble; generalization**

I. INTRODUCTION

Neural network ensemble method can combine training and emulation result of several neural networks, greatly enhancing the generalization ability of the neural network system. Neural network ensemble mainly works on two aspects: create the subsystem of the neural network and integrate the result of the subsystems. The most important techniques of creating the subsystem are Boosting and Bagging[1]. The techniques of integrating the results of the subsystems include: majority voting method, mean method, weighted mean method, evolutionary integration design method based on genetic algorithm 2] etc.

The generalization ability of the neural network system depends on the accuracy and the differentiation of the individual neural networks and usually the same type of supervised learning neural networks are used as individual neural sub network. The method [3] to generate the subsystems of individual neural network is: assigning different initial weight, applying different hierarchical structure, using different training data and algorithms to the neural networks of the same type. In this article, we adopt different types of neural networks with different learning systems (supervised and unsupervised) as subsystems combined with input and target patterns, implementing AHP for pattern integration.

II. BASIC METHODS AND PRINCIPALS

A. SOM neural network

SOM stimulates the characteristic that the brain cells of the same function gather. After the learning process of SOM, the adjacent output units will have the similar function. When neural network accepts external input, the neurons will be divided into different reaction sites, each site have different react characteristics according to the input pattern.SOM neural network is a unsupervised self-organizing network[4] and is composed by input layer and compete layer.

B. PNN

Probability Neural Network (PNN) was proposed by Doctor Specht[5] in 1989, and PNN is a supervised feedforward neural network.

PNN is composed by input layer, pattern layer, summation layer and output layer.

PNN Computing Process:

(1) Construct neural network: the number of input layer neurons is equivalent to the dimension of the input vector, the number of pattern layer neurons is equivalent to the number of training samples and the number of summation layer neurons is equivalent to the number of target patterns. Based on the target pattern of the training sample, we construct the connection between the pattern layer and the summation layer.

(2) Calculate the probability density function of q types of target pattern A^q , $f_{A^q}(x)$: $f_{A^q}(x) = P(x \mid A^q)$, in which x is the input data that is to be classified. The formula is:

$$f_{A^q}(x) = \frac{1}{(2\pi)^{0.5\times P} \cdot \sigma^P} \cdot \frac{1}{n_t} \cdot \sum_{i=1}^{n_t} \exp((x \cdot X_{A_i^q} - 1)/2\sigma^2) \qquad (1)$$

In the formula, P is the dimension of input data; n_t is the number of training sample, $X_{A_i^q}$ is the i^{th} training data of type A^q . σ is the smoothing factor.

(3) Output: $O = \max(f_{A^q}(x))$

C. RBF neural network

RBF[6] (Radial Basis Function) is a neural network structure that was proposed by J.Moody and C.Darken in late 1980s. It is a 3-layer feed forward network which is composed by input, hidden and output layers. The hidden layer uses the radial basis function $f(n)=e^{-n^2}$ as the activation function. The training of RBF network consists of 2 steps: the first is unsupervised learning to determine weight $w1$ between the input and hidden layers; the second is supervised learning to determine weight $w2$ between the hidden and output layers.

(1) Initialize: Assign the value of weight $w1_j(0)$ randomly.

(2) Calculate the similarity: Find out the weight which is the most close to the training sample.

$$d_r(t)=\min(\| x(t)-w1_i(t-1)\|) \tag{2}$$

(3) Adjust the weight.

$$w1_i(t)=w1_i(t-1) \quad 1\le i\le h, i\ne r \tag{3}$$

$$w1_r(t)=w1_r(t-1)+\beta(x(t)-w1_r(t-1)) \quad i=r \tag{4}$$

Out of which, h is the number of neurons of hidden layer and β is the learning rate, $0<\beta<1$.

(4) Let $t=t+1$, return to step (2), compute until $\| w1_i(t)-w1_i(t-1)\|<\varepsilon$

The learning process of second step:

r_i^q is the output of the computed q[th] input pattern by the hidden layer.

(1) Initialize:

(2) Calculate the output of the network's output layer:

$$y^q=\sum_{i=1}^{h} r_i^q \times w2_i \tag{5}$$

(3) Calculate the difference between the result and the target. (s is the number of output neurons.)

$$d_j^q=t_j^q-y_j^q \quad j=1,2,\cdots,s \tag{6}$$

(4) Adjust the weight:

$$w2_{ij}(t+1)=w2_{ij}(t)+\alpha\cdot y_i^q\cdot d_i^q \quad i=1,2,\cdots,h \tag{7}$$

(5) To next training pattern repeat step (2)~(4), until d_j^q is small enough.

D. LVQ neural network[7]

LVQ (Learning Vector Quantization) neural network is supervised training method for compete layer and is composed by input layer, competition layer and the output layer. The input and competition layers are all-connected. self-organized learning and classification of the input vectors are done by the competition layer. The classification result is only related to the Euclidean distance between the input vectors. The competition and output layers are partially connected. Every output neuron is connected to different neuron groups of the compete layer and the link weight is fixed to 1. The input vectors are classified by the competition layer into several subsets and the output layer merges the subsets into the target classification pattern.

Let x^i be the i^{th} training pattern vector, C_{X^i} be the pattern type of the input vector x^i, C_j be the type of the j^{th} output neuron.

The learning algorithm of LVQ network is:

(1) Initialize: Initialize the weight $w_j(0)$ of vector randomly, choose the value of learning rate η, $k=0$.

(2) Check out if the termination condition of the training is satisfied, quit the process if satisfied, or continue if no.

(3) To every training sample, process step (4) and (5).

(4) Determine the wined vector weight w_q, which satisfies:

$$d(x^i,w_q)=\min_{\forall j} d(x^i,w_j)=\min_{\forall j}\| x^i-w_j\|_2^2 \tag{8}$$

(5) Update weight w_q, in which, $0<\eta(k)<1$.

$$\begin{cases} w_q(k+1)=w_q(k)+\eta(k)(x^i-w_q(k)) & C_{w_q}=C_{x^i} \\ w_q(k+1)=w_q(k)-\eta(k)(x^i-w_q(k)) & C_{w_q}\ne C_{x^i} \end{cases} \tag{9}$$

(6) Let $k=k+1$, decreases learning rate $\eta(t)$ and return to step (2).

E. Analytic Hierarchy Process

AHP (Analytic Hierarchy Process) which was established by Professor Thomas Saaty guarantees the scientificity of the qualitative analysis and the accuracy of the quantitative analysis. To solve decision-making problem, factors of multiple aspects and the judgement criteria must be considered and decisions are made via these criteria. AHP arranges and stratifies the problem. AHP creates a 3-layer structural model[8] as shown in Figure3.

(1) Target Layer: The only target or expected result of the decision making problem.

(2) Criteria Layer: Include all the needed criteria and the sub-criteria to achieve the target.

(3) Measure Layer: Include the different available methods.

Pairwise comparison method can determine the effect of every factor to the same target. Suppose $y_1, y_2, \cdots y_n$ are the influences of n factors to the same target, 2 factors y_i and y_j are chosen each time, a_{ij} is the ratio of the y_i and y_j; and $a_{ji}=1/a_{ij}$. In which, a_{ij} can be found in Table1.

TABLE I a_{ij} V alues

Value	Meaning	Value	Meaning
1	Equally	7	Remarkably more important
3	Slightly more important	9	Extremely More important
5	Obviously more important	2,4,6,8	Mean value

Definition 1:If matrix $A=(a_{ij})_{n\times n}$ satisfies $a_{ij}>0$, $a_{ji}=1/a_{ij}$ $(i,j=1,2,3\cdots n)$, then A is a reciprocal matrix.

Definition 2: If reciprocal matrix A satisfies the equation: $a_{ij}\cdot a_{jk}=a_{ik}$ $\forall i,j,k=1,2\cdots n$, then A is a consistent matrix.

The process of consistency checking:

(1) Calculate Consistency Index, CI, $CI=(\lambda_{\max}-n)/(n-1)$.

(2) Find out Random Index, RI is shown in the table below.

TABLE II Random Index

n	1	2	3	4	5	6	7
RI	0	0	0.52	0.89	1.12	1.26	1.36
n	8	9	10	11	12	13	14
RI	1.41	1.46	1.49	1.52	1.54	1.56	1.58

(3) Calculate CR (Consistency Ratio), $CR=CI/RI$. If $CR<0.1$, the consistency of the matrix is acceptable, otherwise, the matrix must be modified.

(4) Processtotal ranking, consistency checking.

(5) The upper layer A includes m factors, $A_1, A_2, \cdots A_m$, the total ranking weights are $a_1, a_2, \cdots a_m$ am respectively. The next

layer (layer B) includes n factors, $B_1, B_2, \cdots B_n$, their single ranking weights according to A_i are $b_{1j}, b_{2j}, \cdots b_{nj}$ respectively. In layer B, the total target weights $b_1, b_2, \cdots b_n$ for each factor. The computing formula is:

$$b_i = \sum_{j=1}^{m} b_{ij} \cdot a_j \quad i = 1, 2, \cdots, n \tag{10}$$

The CR of layer B total ranking is:

$$CR = (\sum_{j=1}^{m} CI(j) \cdot a_j) / (\sum_{j=1}^{m} RI(j) \cdot a_j) \tag{11}$$

If $CR < 0.1$, the consistency of the matrix is acceptable, otherwise, the matrix should be modified.

The measure with the maximum total weight will be adopted as the decision.

III. INTEGRATED SYSTEM MODEL

The ensemble of the fault pattern recognition can be seen as a decision-making process which is based on analyzing the recognition results of the input vectors and pattern recognition subsystems and chooses among the typical fault patterns. Compute the measurements from input vectors to typical fault patterns, for example, the number of recognized typical faults by the subsystem, the expectations and variances of input vectors and typical faults, the degree of membership of fuzzy classification and the roughness of the rough set. By paired comparison the measurements, the matrix is established. Analyze the paired comparison matrix using AHP and obtain the ultimate pattern recognition result.

It is easier to integrate the system when the differentiation is larger among subsystems.To make the subsystems mutual independent and the pattern recognition irrelevant, we use supervised and unsupervised learning neural networks to identify the subsystems. AHP method is used to integrate the recognition pattern of each subsystem after pattern mapping and standardization into final recognition pattern. The frame of the integrated system is shown in Figure.1

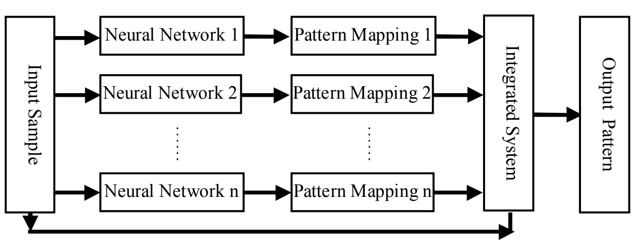

Figure 1. Frame of integrated system

The recognition result of supervised learning neural network is unique while the result of unsupervised learning is indefinite. After the training of neural network, by using typical errors, we virtualized the unsupervised learning subsystem and obtained the mapping: $y : I_t \rightarrow O_j^{(i)}$ in which I_t is the typical error input, $O_j^{(i)}$ $(i = 1, 2 \cdots m, j = 1, 2, \cdots, k)$ is the output of subsystems. The mapping y is one to one correspondence and its inverse mapping is y^{-1}. The recognition outputs after the pattern mappings of subsystems are: $y^{-1}(x) = I_t$ which $I_t \rightarrow O_j^{(i)}$ and satisfy:

$$\| x - O_j^{(i)} \| = \min(\| x - O_z^{(i)} \|) \quad z = 1, 2, \cdots k \tag{12}$$

A. *The integrated model of hierarchical analysis*

The number of recognition pattern of neural network subsystem can reflect the pattern of input vector to some extent. The expectation and square deviation of the difference of inputand target vectors can also reflect the pattern of input vector.

The ensemble model of hierarchical analysis is shown in Figure 2. Measure layer is the pattern that requires recognition. Criteria layer is the number of recognition patterns plus the expectation and square deviation of the difference of input and target vectors.

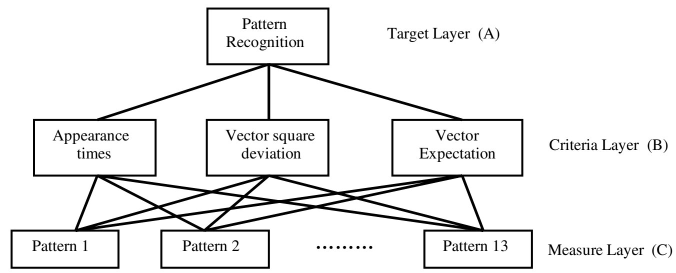

Figure 2. AHP Model.

The Process of the ensemble computing:

(1) Create each neural network subsystem;train each subsystem independently and calculate the pattern mapping of subsystem. Apply pattern recognition on input pattern vectors and count the number of recognition patterns.

(2) Compute the expectation and square deviation of the difference of input and target vectors.

(3) Construct the pair-wise comparison matrix from the number of the pattern, the expectation and square deviation of the difference vector respectively.

(4) Compute and normalize the eigenvector with the maximum eigenvalue, check the consistency of each criterion.

(5) Compute the total weight vector of measure layer pattern to target layer pattern and check the total consistency; obtain the final recognition pattern according to the total ranking weight vector.

B Performance Analysis:

1. The number of the subsystems in the ensemble of AHP can be odd and even. The problem of the relative majority method that the number of maximum appearances is shared by two patterns can be solved while using AHP.
2. The relative majority method treats each pattern recognition subsystem equally. However, the experiment shows that different pattern recognition subsystems have different recognition rates, that is to say each subsystem contributes differently to the final result. AHP determines the weight of each subsystem based on their recognition rates and the ultimate fault recognition rate can be improved.

IV. EXPERIMENT

Using the integrated system to train and emulate typical fault samples of the specific equipments.

TABLE III TYPICAL FAULT SAMPLES

	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8
y_1	0.25	0.60	0.25	0.50	0.50	0.25	0.50	0.25
y_2	0.25	0.60	0.5	0.50	0.50	0.75	0.50	0.50
y_3	0.75	0.40	0.75	0.50	0.50	0.50	0.50	0.50
y_4	0.25	0.60	0.50	0.25	0.25	0.75	0.50	0.50
y_5	0.25	0.60	0.50	0.00	0.00	1.00	0.50	0.50
y_6	0.75	0.40	0.75	0.25	0.25	0.75	1.00	0.25
y_7	0.25	0.60	0.50	0.50	0.50	0.75	0.75	0.50
y_8	0.75	0.40	0.50	0.75	0.75	0.25	0.00	1.00
y_9	0.25	0.60	0.50	0.75	0.25	0.25	0.50	0.50
y_{10}	0.25	0.60	0.50	0.25	0.25	0.75	1.00	0.50
y_{11}	0.50	0.50	0.50	0.75	0.75	0.25	1.00	0.50
y_{12}	0.00	0.75	0.75	1.00	0.75	0.25	1.00	0.50
y_{13}	0.50	0.50	0.50	0.75	0.75	0.25	0.50	0.50

（X is the sign set, x_i (i=1,2,…, 8),Y is the fault set y_j （j=1,2,…13)。）

Adding random noise signals with square obviation values as 0.050.080.120.20.30.40.5 into typical faulty data, we obtain 7 groups of testing data, each contains 100 faults. Testing data are 8*1300 matrixes. Using 4 different types of neural networks to construct 5 recognition subsystems: the first subsystem is SOM, with 5 *5 array as its output layer and training samples as typical faulty data trained 100 times. The second subsystem is PNN, with training samples as 13 typical faulty data trained 100 times and Radial Basis Functions having 0.1 distribution density. The third and fourth subsystems are two LVQ neural networks with different parameters, with SPREAD as 1.0 and 20, the number of neurons in hidden layer as 30 and 20 respectively. Training samples are 13 typical faulty data trained 300 times. The fifth subsystem is RBF, with training sum-squared error target as 0.01, SPREAD as 40 and training samples are 130 faults in which each type of fault appears exactly 10 times. The mean value of correct recognition rate of the pattern is obtained by computing each group of testing data 10 times. In AHP, based on target layer, criteria layer has weights as 0.3, 0.3, 0.4. When the difference of the number of patterns is 0, 1, 2, 3, 4, 5, respectively, the pair-wise comparison is $a_{ij} = 1,2,3,4,5,6$. Suppose 7e is the expectation of the difference of the input vector and 13 target vectors; the possible difference of the maximum variance and the minimum variance could be $0, e, 2e, 3e, 4e, 5e, 6e, 7e$; the pairwise comparisons are $b_{ij}, c_{ij} = 1,2,3,4,5,6,7,8$ respectively. The classification correct rate comparison chart is shown in Figure 3.

The experiment indicates:

(1) The recognition rate of integrated system is higher than each subsystem and the rate increment is high in low noise condition.

(2) RBF and SOM subnets have higher recognition rates in low noise condition compared to other subnets, while in other conditions the rates are much lower.

(3) LVQ and PNN subnets with different initial parameters have similar recognition rates

(4) AHP has a higher recognition rate than Relative Majority Integrated Method, with remarkably better effect in medium and low noise conditions. While in high noise condition, the two methods have similar recognition rates.

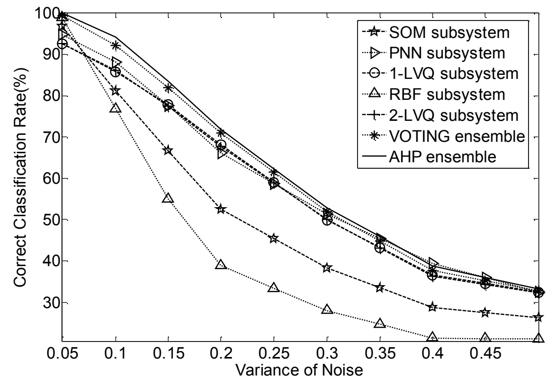

Figure 3. The classification correct rate comparison chart.

V. CONCLUSION

This article establishes the AHP ensemble model and uses typical faulty data of specific equipments for testing experiments. We adopt 5 neural networks as subsystems in 4 types, SOM, PNN, 1-LVQ, 2-LVQ and RBF. Train each subnet independently and increase pattern recognition rate by using AHP to integrate the recognition pattern of neural network subsystems. The integration method consumes little computing resource and the result of calculation accords with the actual conditions, which indicates that AHP is an efficient ensemble method.

REFERENCES

[1] Zhou Zhihua, Chen shifu. Neural Network Ensemble[J], Chinese Journal of Computers , 2002.1, Vol.25, No.1.

[2] Jared Sylvester, Nitesh V. Chawla..Evolutionary Ensemble Creation and Thinning[J], Neural Networks, 2006.7 16-21:5148 - 5155.

[3] Sharkey AJC.On combining artificial neural nets[J], Connection Science, 1996,8(3-4):299-314.

[4] Kohonen.The Self-Organizing Map[J],Proceedings of IEEE, 1990,78(9):1464-1481.

[5] Specht D F.Probabilistic neural networks[J], Neural Networks, 1990, 1(3): 109-118.

[6] Li Guoyou, Yao Lei.TA New Method of Pattern Recognition Based on Optimized RBF Neural Networks[J], Journal of System Simulation, 2006.1,Vol.18 No.1:181-184

[7] Lou Suntian.System Analysis and Design Based on Matlab-Neural Network[M], XiDian University Press , 1999.

[8] Saaty T L.The Analytic Hierarchy Process: Planning,Priority setting, resource allocation[M], McGraw Hill, , 1980.

[9] Wang Yan, Chen Xiang-dong, Jiang Ya-dong, etc.Study on electronicnose recognitionmethod with independent component analysis and backpropagation neuralnetwor[J], Transducer and Microsystem Technologies, 2006,25(9):36-39 .

[10] Lu Shuang, Hou Yue-qian.FaultPattern Recognition of Rolling Bearing Based on PrincipalComponentsAnalysis and RadialBasis Function NeuralNetwork[J], Maching Tool and Hydraulics, , 2005,3:185-188.

[11] Sven Nordholm ,Siow Yong Low.Speech Signal Extrac- tion Utilizing PCA-ICA Algorithm whith Non-Uniform Spacing Microphong Array[J], ICASSP 2006 V 965-968.