International Journal of
Intelligent Engineering & Systems

http://www.inass.org/

Rough Sets -Least square and Neural Networks in Fault Diagnosis Shield Applied Research

Yang Yu*, **Chao Han**

School of Information Science & Engineering, Shenyang Ligong University
Shenyang, China

* *Corresponding author's E-mail:yusongh@126.com*

Abstract: Shield is a typical mechanical, electrical, hydraulic integration of equipment. Its faults species are complex and diverse. To prevent because machine failure causes economic losses and casualties by shield, this article will introduces rough set theory to the subway shield machine fault diagnosis, propose a method which is based on rough set theory combined with neural network of Metro shield machine fault diagnosis. Use the strong advantage of rough sets theory in attribute reduction, and remove the data redundancy of information which is not effective for decision-making. Application of neural network algorithm to reduce date for diagnosis, the method can effectively improve the speed and accuracy of the diagnosis. Then use BP neural network combined with least square method to forecast fault, the least-square can reflect the trend of linear sequence, Neural network can seize the variation of nonlinear time series, therefore the combination of two methods could well predict the future of unit operating conditions.

Keywords: Rough set; Neural network; Shield machine; Fault diagnosis; Least square

1. Introduction

Shield machine is the most important equipment in the process of mining [1, 2]. Once occur fault not only cause huge economic losses, but also may cause casualties. Shield machine is a mechanical, electrical, fluid integration of equipment. Failure has the features of complexity, uncertainty and diversity, etc. And in the process of operation that needs to be monitored data is very large, redundant data is more also. Therefore it brings great difficulties to fault diagnosis. The traditional expert system and neural network fault diagnosis methods have the problems of knowledge acquisition bottleneck, network structure complex, long training time issues. The diagnostic efficiency is unsatisfactory, not conducive to real-time diagnosis of the Shield machine. In addition, Monitoring system collect a lot of running status attributes of the equipment. But not all of the features data have influence to the fault generate. Rough set theory [3, 4, 5] in the data attribute reduction is very effective. And the neural network classification ability is very strong. So this paper proposes a method which is based on rough set combine with neural network fault diagnosis and uses BP neural network combined with least square method to forecast fault.

2. Rough set theory

Rough set theory is proposed by Warsaw University of Technology Z.Pawlak that deals with incomplete, imprecise knowledge representation, learning, inductive method. The theory does not require any additional prior knowledge. For example, statistics required prior probability, fuzzy sets required membership functions and so on. The research of theory and applications has become a hot spot at home and abroad. The application of rough sets theory for knowledge discovery, data reduction, decision support, classification, pattern recognition and others has proved to be a very effective new mathematical approach. Rough set theory that "let the data speak for itself" is very convenient for the idea of its application in fault diagnosis field.

2.1 Knowledge representation system

In rough set theory a knowledge representation system S can be expressed as $S =< U,C,D,V,f >$, in which U is a collection of objects, $C \bigcup D = A$ is Attribute Set, Subset C and D respectively called condition attribute and decision attribute. $V = \bigcup_{a \in A} V_a$ is attribute values collection. V_a express the scope of the attribute $a \in A$. $f : U \times A \rightarrow V$ is a Information function. It specifies the U of each object x attribute values.

2.2 Knowledge reduction and core

Given universe of discourse U, order R is a family of equivalence relations which is over the universe of discourse U and $r \in R$, when $ind(R) = ind(R-r)$, called r can be dispensable for the R. Otherwise called r can not be dispensable for the R. As for any $r \in R$, if R is indispensable, then family R is independent. Given universe of discourse U, Set P, Q are two equivalence relation families which defined on U. when Q is independent, and $ind(Q) = ind(P)$, $Q \subseteq P$, called Q is a reduction of P. Obviously P can have many kinds of reductions. All the indispensable collection of relations among P, is called $P's$ core, denoted by Core(P). In which $core(P) = \bigcap(P)$, $red(P)$ is all the reduction families among P. There are two aspects about the reduction, which can see from the above. First, expression by the reduction division of the system with the original division of the knowledge base is fully consistent. Expression by the reduction of knowledge and knowledge of the original has the same expressive power. Second, that is independence. Reduction inside non-further reduction, and family of equivalence relations $R's$ reduction is not unique.

Definition 1: set $X \subseteq U$, R is the equivalence relation on U There $R_*(X) = \bigcup Y \subseteq U/R : Y \subseteq X$, $R^*(X) = \bigcup Y \in U/R : Y \subseteq X$ called $R_*(X)$ is X Lower Approximation, $R^*(X)$ is X Upper approximation. Y is one equivalence relation of U. $POS_R(X) = R_*(X)$ called $X's$ R positive region, positive region $POS_R(X)$ is a set of all certainly classified as elements of the set of X which is based on knowledge of R,U.

Definition 2: Set up $DT =< U,C,D,V,f >$ is one decision table. The condition attributes and decision attributes are respectively C and D. Said D in DT with the extent $k(0 \leq k \leq 1)$ dependent on C. Among $k = r_C(D)\frac{card(POS_C(D))}{card(U)}$, (D) is $D's$ C positive region. Obvious K is the knowledge of C about $D's$ approximate quality. It reflects the knowledge of C to D of the classification ability. In other words, dependence is another kind of classification. If $k = 1$, D is totally dependent on C. If $0 < k < 1$, D partly depends on C. If $k = 0$, D does not depend on C.

Definition 3: Set up $DT =< U,C,D,V,f >$, For C non-empty subset B. The degree of importance is $\delta_{CD}(B) = \gamma_C(D) - \gamma_{C-B}(D)$. As can be seen from the definition, if the $B's$ importance is 0, Said B can be removed from the C, that B is redundant.

Definition 4: Set $\varphi \rightarrow \psi$ is a decision rule on decision table. V is an attribute value which can be about to. If and only if $(\varphi \rightarrow \psi) \rightarrow (\varphi - (av) \rightarrow \psi)$, among φ and ψ are the decision table logic formula. This theorem reveals that a decision rule's condition attribute values may be about to. If and only if after about to, that remains the consistency of this rules.

Definition 5: Function d_X: $A \rightarrow V$, make $d_X(a) = a(x)$, among, $a \in A, X \subseteq U$, $x \in U$, called d_X is a decision rule of decision table $DT =< U,A,V,f >$. If $a \in C \subset A$, denoted $d_X|C$ called condition part of decision rules. decision rules.

3. Neural network theory

Artificial Neural Network-ANN model is based on modern neurophysiology and psychology, mimicing the structural characteristics of brain neurons established by a nonlinear dynamical system, which consists of a large number of simple nonlinear processing units highly parallel, interconnect becomes, and has some capabilities of human brain with certain basic features of simple mathematical modeling, therefore, from the purpose to be achieved, neural network and what is known as artificial intelligence technology are the same, but both achieving the same purpose is not the same train of thought. Artificial intelligence focuses on the "Functional Simulation", but the neural network through the simulation of human brain structures is to realize the function of the brain information processing simulation. Application of this simulation is to solve engineering problems.

3.1 The basic principles of BP network

The basic principles of BP network, BP neural network (Back Propagation Neural Network), also known as back propagation neural network, by the nonlin-

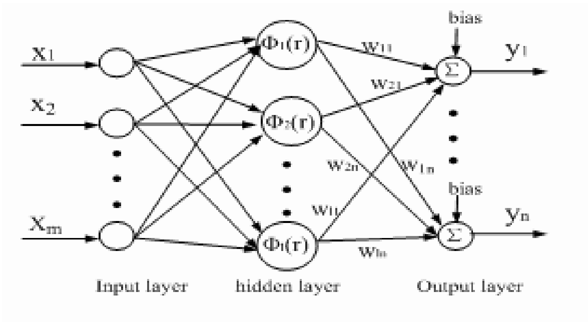

Figure 1. BP network topology

ear transformation unit composed of feedforward network. Among many neural network models, BP network is a model which is the recognition of the earliest and most widely used. In practical applications of artificial neural networks, 80%-90% use *BP* network, or its variations. Now, BP network in pattern recognition, image processing and analysis, control and other fields has a wide range of applications. In the field of fault diagnosis is the most successful application of a network model.

3.2 BP neural network model

BP network topology as shown in Figure1.

The BP neural network's features are: Network including the input and output layer, has one or more layers of hidden layer. The same level there is no coupling between the nodes; Input signal from the input layer nodes, in turn passing through the hidden layer nodes. Finally reaching the output layer nodes, each node of the output layer only influences the next output layer nodes; each node of the cell characteristics (activation function) is usually S-type function. In the output layer, the node activation function is linear at times.

BP neural network is a multilayer feedforward network. In this paper, three BP neural network models are used, which consist of input layer, hidden layer and output layer.

(1) Input layer node i, whose output equals input $x_i(i = 1,2\ldots n)$ to control variable transmission to the second layer.

(2) Hidden node j, input h_j, output y_j, as follows:

$$h_j = \sum_{i=1}^{n} \omega_{ij}\chi_i - \theta_j = \sum_{i=1}^{n+1} \omega_{ij}\chi_i \tag{1}$$

$$y_j = f(h_j) = \frac{1}{1+e^{-h_j}} \tag{2}$$

Among $j = 1,2\ldots m; \theta_j = w_{n+1/j}, x_{n+1} = -1$

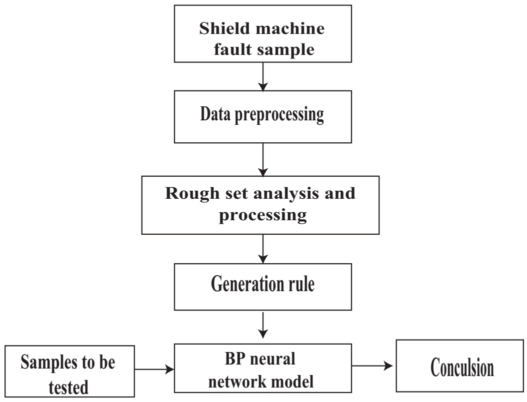

Figure 2. Rough Set and Neural Network Fault Diagnosis Model of Shield machine

(3) Output node k, input h_k, output y_k as follows:

$$h_k = \sum_{i=1}^{m} \omega_{jk}o_j - \theta_j = \sum_{i=1}^{m+1} \omega_{jk}o_j \tag{3}$$

$$y_k = f(h_j) = \frac{1}{1+e^{-h_k}} \tag{4}$$

Among, $k = 1,2\ldots r$; $\theta_j = w_{(m+1)}/o_{m+1}$, $o_{m+1} = -1$.

When given an input mode of network, it transforms from the input layer to the hidden layer unit cell. Through the hidden layer post-treatment, then sent to the output layer, processed by the output layer unit, it produces an output mode. This is a layer status update process, called forward propagation. If the error between output response and the desired output mode does not meet the requirements, then into the error back propagation. The error along the connection path layer transmission layer by layer modifies every layer's connection weight, until meets the requirements.

4. Fault diagnosis process

Using rough sets theory combined with BP neural network fault diagnosis method is to regard rough set as a pre-system, as to the fault sample pretreatment before fault diagnosis (That is discretization), and then proceed to attribute reduction, eliminate redundant attributes and repeat attributes, and then use the reduced decision table as BP neural network input, proceed fault diagnosis. Shield machine fault diagnosis model shown in Figure 2.

5. Application analysis

This paper uses the North Heavy Industry Group (Shenyang) productive slurry balance shield machine common errors as object and combines it with the

Table 1. The decision table of shield machine fault diagnosis

Fault	Condition attribute									Desiccant
Sample	$C1$	$C2$	$C3$	$C4$	$C5$	$C6$	$C7$	$C8$	$C9$	attribute
$U1$	3.1	43.4	38.1	396.5	381.2	21.6	335.6	42.4	37.3	1
$U2$	2.9	40.0	38.2	396.1	380.3	20.8	34.4	38.1	36.9	1
$U3$	3.0	39.5	40.5	395.4	379.4	21.8	36.7	41.2	37.4	2
$U4$	2.9	42.3	38.0	398.6	381.5	22.7	35.9	40.8	39.9	3
$U5$	3.2	43.0	36.7	394.2	380.1	19.7	34.6	40.3	43.0	4
$U6$	3.0	38.6	38.6	396.3	378.3	20.5	35.0	39.6	40.7	5
$U7$	2.9	39.0	37.7	393.2	381.1	21.8	33.5	37.9	42.1	6

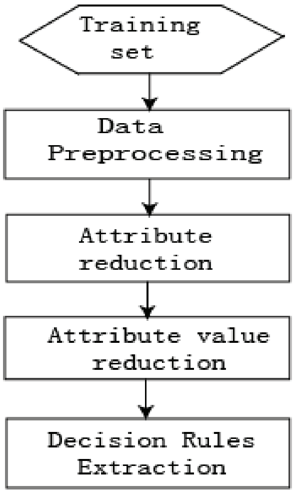

Figure 3. Decision table data reduction process

thesis [7] to classify the dynamic real time measuring project of subway shield machine. Select data acquisition system to keep the parameters which are related with the cutter: Grouting pressure(3), Grouting pressure(4), $No.1$ Cutter clutch temperature, $No.2$ Cutter clutch temperature, $No.3$ Cutter clutch temperature, $No.4$ Cutter clutch temperature, $No.5$ Cutter clutch temperature, $No.3$ power supply voltage, $No.4$ power supply voltage, Nine parameters as the condition attribute, Motor valve damage, Individual drive does not move, Clutch not excitation, Individual motors don't turn, Cutter clutch overheating, Supply lipid PU which starts failure as Decision attribute(with 1, 2, 3, 4, 5, 6 express). Establish fault diagnosis decision table shown in Table 1.

5.1 Data preprocessing

When using Rough sets theory of decision table attribute reduction, the process as shown in Figure 3.

Use rough set theory for data mining, needing to do some necessary process to the collected data. Data preprocessing includes data preparation, handling missing values, and data discretization. Because shield machine data acquisition system is very integrated, only discretization of data is considered. Data discretization methods [8] mainly include:

(1) Equidistant partition algorithm:

Such discretization algorithm is in every attribute, according to the parameters given by the user to the attribute values into a simple paragraph break point, equidistant does not consider each breakpoints paragraph's number of attribute values. Suppose the maximum property value of $Xmax$, minimum property value of $Xmin$, the user given Parameters is k, The breakpoint interval is $n = (Xmax - Xmin)/k$. Get this breakpoint on the attribute is $Xmin + i \times n$, $i = 0, \ldots, k$. Equal to the distance between these breakpoints.

(2) Such frequency partition algorithm:

Such discretization algorithm is based on the parameters given by the user k to m objects into paragraph, Each section has m/k objects. Suppose the maximum property value of an attribute is $Xmax$, minimum property value is $Xmin$, the user given Parameters is k, need to be the property's value in all instances be arranged from small to large, then average divided into k segment get break point set. Between two adjacent breakpoints each attribute value contains the equal number.

(3) Naive Scaler algorithm: For each attribute $a \in c$, the following procedure:

According to the value of $a(x)$, for example $x \in u$ from small to large order; Scanned from top to bottom, Set X_i and X_j represent two adjacent instances, if $a(x_i) = a(x_j)$, continue to scan, if $d(x_i) = d(x_j)$, the same decision continues to scan, else get a breakpoint $c = (a(x_i) + a(x_j)/2)$.

According to the fault samples, the properties of a single failure have no obvious change. So use equidistant partition method, $N = 0, 1, 2 \ldots\ldots$express attribute values of interval after discretization. For example: $C1$ Division interval is [2.9,2.975), [2.975, 3.05), [3.05, 3.125), [3.125, 3.2]. Other property values are as this division. Decision table after discretization as shown in Table 2.

5.2 Rough set attribute reduction

This paper uses a kind of method based on Feature selection (Attribute significance). From the $core(P) = \bigcap red(P)$ known, any decision table's relative core is unique. It contains all relative reductions, according to the definition 2 to the definition of attribute importance. From the original condition attributes subset

Table 2. Decision table after discretization

Fault Sample	C1	C2	C3	C4	C5	C6	C7	C8	C9	Desiccant attribute
	Condition attribute									
$U1$	2	3	1	2	3	2	2	3	0	1
$U2$	0	1	1	2	2	1	1	0	0	1
$U3$	1	0	3	1	2	2	3	2	0	2
$U4$	0	3	1	3	3	3	2	2	1	3
$U5$	3	3	0	0	2	0	1	2	3	4
$U6$	1	0	2	2	1	1	1	1	2	5
$U7$	0	0	1	2	3	2	0	0	1	6

Table 3. Minimal decision table

Fault samples	Condition attribute				Decision attribute
	$C3$	$C4$	$C8$	$C9$	
$U1$	1	2	3	0	1
$U2$	1	2	0	0	1
$U3$	3	1	2	0	2
$U4$	1	3	2	1	3
$U5$	0	0	2	3	4
$U6$	2	2	1	2	5
$U7$	1	2	0	3	6

(Suppose there are K attributes) delete attribute significance smallest property operator, get $K-1$ condition attribute's sub-attribute decision table, and then repeat the operation in these Sub-decision table. And to $\gamma_C(D)$ as a reference, it eventually can get a relative reduction. That's the core attribute. It plays a decisive role in the decision table. Realization process is as follows:

Input: decision table $DT = < U, A, V, f >, A = C \bigcup D$, $C = c_1, c_2, \ldots, c_k$, Calculated D on C dependence $\gamma_C(D)$, set $Red = C(Redisareduction)$, While $(\gamma_{Red}(D) == \gamma_C(D))$, Calculating $Red's$ all attributes importance, according to the importance of all the properties on the Red sorting. Select the properties of the smallest importance $a \in Red$. IF $(\gamma_{Red-\{a\}}(D) == \gamma_C(D))$ $\{Red = Red - \{a\}\}$, Else $\{PRINT(Red)\}$. For information after attribute reduction table, not all the conditions of each record attribute values are required. Value reduction is investigated in table records, delete all the rules does not affect the expression of the redundant condition attribute values, and remove the decision table's rows redundant column values. Information table after reduction has less number of attributes and less attribute values than the original information table. After value reduction it can be automatically extract the rules. But each case may have multiple value reductions which are the same as attribute reduction. In practice, not all the value reduction can be obtained, just concerned about the combination of minimal value reduction (After value reduction the different examples are least) [8].

After reduction treatment [9], receiving one of the smallest reduction is $\{C3, C4, C8, C9\}$. Reduced the minimal decision table as shown in Table 3.

5.3 Neural network Diagnosis

5.3.1 BP neural network design

When designing the BP neural network, general from the network layer, the number of neurons in each layer and the activation function, learning rate of the initial value and the aspects are to be considered. Theory has been proved: A three-tier BP network can approximate any precision to a nonlinear function mapping. This fact has given us a basic design principle of BP network. Increasing the number of layers can further reduce the error, improve the accuracy, but also make the network more complicated, thus increasing the network's training time. In fact improving the precision error can also use a hidden layer to increase the number of neurons to obtain, which is in structure to achieve more simply than increasing the hidden layer, and its training effects are easier to observe than the increase in number of layers and adjustment. Therefore, under normal circumstances, priority should increase the number of neurons in hidden layer. The selection of hidden nodes: First use of general empirical formula:$k = \sqrt{n+m} + a$.Among, (m is the number of output neurons, n is the number of input neurons, a is a constant between 1 to 10) roughly estimate the range, when specific design, through training different number of neurons to compare to find the optimal number of hidden layer units.

Summarily, the training function apply traindx, Learning function apply learngdm, Performancece function apply mse, Select S-logarithmic (logsig) and the linear function (purelin) as the transfer function of neurons. The BP neural network performance curve as shown in Figure 4, Figure 5.

From training error and training result diagram can be seen, the classification results are basically the same, however, through reduction of rough set algorithm of fault sample after training the neural network shorter training time, fewer training steps, more accurate. As shown in table 4.

This paper adopts threshold discriminant conditions to fault diagnosis. Set the threshold of 0.5. Give a set of failure data input. If the output threshold is less than 0.5, we believe that the failure does not oc-

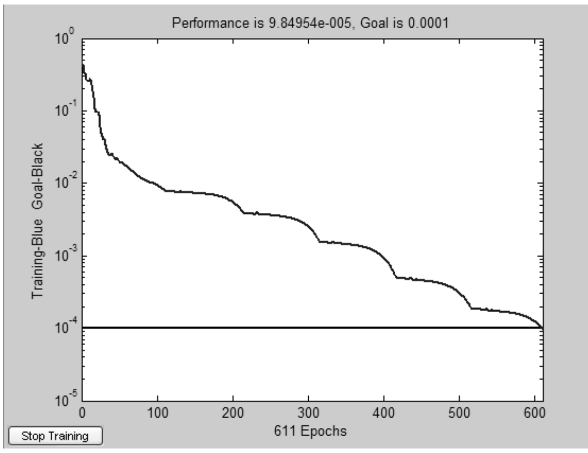

Figure 4. The neural network training error before the attribute reduction

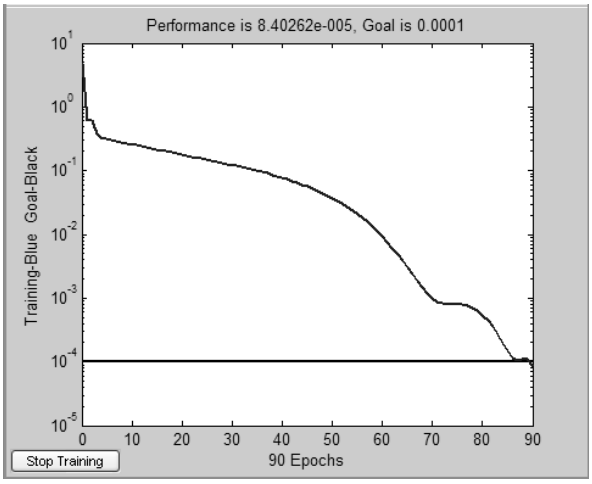

Figure 5. The neural network training error after the attribute reduction

Table 4. Performance Comparison of BP neural network

Fault samples	Training time	Training times	MSE
Unreducted Attribute	3.5158	611	9.84954e-005/0.0001
Reducted Attribute	1.2614	90	8.40262e-005/0.0001

cur. If more than 0.5, the failure occurres. Using the original fault samples and through the rough sets attribute reduction of fault samples respectively train the BP neural network. After determining the parameters of two neural network, put table 5 pre-diagnosis fault samples into two neural network for diagnosis. Fault samples to be tested. The Diagnosis result as shown in Table 6.

As can be seen from Table 5, BP neural network and Rough set BP neural network diagnosis results are all 4 faults(Clutch not excitation) output threshold value

Table 5. Fault samples to be tested

$C1$	$C2$	$C3$	$C4$	$C5$	$C6$	$C7$	$C8$	$C9$
2.9	43.0	39.5	396.4	377.9	20.6	35.7	40.8	36.9

Table 6. Comparison of diagnosis results for two neural network

BP neural network	Roughset+ BP neural network	Fault type
0.0468	0.1035	1
0.1534	0.1642	2
0.1236	0.2435	3
0.7825	0.9536	4
0.0675	0.0954	5
0.1428	0.2583	6

greater than the set threshold, respectively 0.7825 and 0.9536. Other kinds of output threshold is far less than the set value. Therefore what can be drawn from the fault is Clutch not excitation. Conventional BP neural network is for fault diagnosis. The enter number of neurons is 9. The fault of Clutch not excitation's threshold is 28.25higher than the set threshold, but this paper uses method which only needs 4 neurons and the fault of Clutch not excitation's threshold is 45.36 higher than the set threshold. The diagnosis result is more accurate.

6. Fault prediction

From the concept of fault diagnosis we can know, failure prediction is essentially a part of fault diagnosis, and the ultimate goal of the diagnosis is to guide the operation and maintenance, but the forecast is to predict the future development trend of unit. Determining whether to adjust or change its operation mode, in fact, comparing with the fault diagnosis, state forecast is a feature of more urgent on-site production. Because the unit will not always be broken, scene more concerned is what is the state when the units continue to run down, in order to better guide the production. At present, the common trend forecasting methods are:

(1) Potential diagram analysis method.

This method is widely used in the operation department, which is by means of machine history data trends, hoping that analysis of the future trend can grasp the moment of the running unit.

(2) Curve fitting method.

First, curve fitting a curve parameter equation, then according to system response time series of historical data and initial value, using Smooth extrapolation techniques to continue curve fitting, according to the curve can be predicted.

(3) Time series method.

Time series forecasting method is commonly referred to as time series analysis method. The basic idea of this approach is the use of autoregressive moving average model, according to the time series model calculating all kinds parameters of the numerical solution, and then under the model prediction.

(4) Prediction Method Based on Grey Theory.

(5) Artificial neural network forecasting method.

In recent years, BP artificial neural network with its unique network of association, memory, storage and learning capabilities and highly nonlinear mapping capability has been widespread concerned in the equipment fault diagnosis, fault prediction. Currently, it has been reported in the literature of neural network models in different fields of extensive research and application, such as the exchange rate forecasting, revenue forecasting, precipitation forecasting, load forecasting, prediction of mechanical condition etc. Therefore Neural network in fault prediction and diagnosis with a good prospect [11].

But for subway shield machine system usually appearers complex and uncertain behavior, a single BP neural network model can not reflect the true time series variation. Shield systems often have a certain trend stationarity, It adopts least-square fitting which can reflect the general trend of change. On this basis, this article adopts the least square method combined with the neural networks for further prediction. The least-square can reflect the linear sequence trend towards, neural networks can seize the variation of nonlinear time series, complementary, which could be a good choice.

For the time series,$X = \{X_t|t = 0,1,2\ldots m\}$, Time series prediction, that is based on sequence$\{x_k,x_{k-1},\ldots xk-n+1\}$Strike x_{k+1}. $x_{k+1} = f(x_k,x_{k-1},\ldots x_{k-n+1}) = f(X_k), n-1 \leq k \leq m$,Type of $X_k = f(x_k,x_{k-1},\ldots,x_{k-n+1})^T$,By the $f(X_k)$complexity, can be divided into linear and nonlinear two parts:

$$x_{k+1} = f(X_k) = L(X_k) + N(X_k) \tag{5.1a}$$

$$L(X_k) = \sum_{i=0}^{n-1} a_i x_{k-1} = {X_K}^T A \tag{5.1b}$$

$$A = (a_0,a_1,\ldots a_{n-1})^T$$

$$N(X_k) = \sum_{i=1}^{P} \beta_j f(\sum i = 0^{n-1} \omega_{ij} x_{k-i} + \theta_j \tag{5.1c}$$

Among, P is the number of hidden layer neurons. (5.1) the parameters in the calculation can be divided into two parts. First, use the least squares method to obtain the $L(X_k)$ parameters m, Then under the least squares residuals obtain the $L(X_k)$ parameters β_j, ω_{ij}, θ_j .Use the least squares residuals to obtain the $L(X_k)$ parameters a_j , By the type (5.1b) can see

$$e_{k+1} = x_{k+1} - {X_k}^T A \tag{5.2}$$

$$\begin{bmatrix} e_m \\ e_{m-1} \\ \vdots \\ e_n \end{bmatrix} = \begin{bmatrix} x_m \\ x_{m-1} \\ \vdots \\ x_n \end{bmatrix} - \begin{bmatrix} x_{m-1}^T \\ x_{m-2}^T \\ \vdots \\ x_{n-1}^T \end{bmatrix} A \tag{5.3}$$

$$E = Y - {X_k}^T A \tag{5.4}$$

Type of $E = (e_m, e_{m-1}, \ldots e_n)^T \in R^{m-n}$

$Y = (x_m, x_{m-1}, \ldots x_n)^T \in R^{m-n}$

$X = (X_m - 1^T, X_m - 2^T, \ldots X_n - 1^T)^T \in R^{n \times m-n}$

According to least squares, A parameter can be obtained, meeting

$$\|E\|^2 = \|Y - X^T A\|^2 = min \tag{5.5}$$

After strike A parameters, substituting (5.1a), available

$$N(X_k) = x_{k+1} - L(X_k) = x_{k+1} - {X_k}^T A \tag{5.6}$$

According to (5.6) available neural network output and input samples:

$\{(N(X_{n-1},X_{n-1})),(N(X_n),X_n),\ldots(N(X_m),X_m)\}$

From this sample, can be obtained by the BP neural network algorithm parameters. When the input is, , desired output is k=n-1,nm, use this sample set as training sample set, and use the BP algorithm to train network. Shield machine running the vibration signal has been proven as non-stationary random process, use run-time vibration signals to failure prediction is an effective prediction method. Select 2009.10.5-2009. 10.6 two days operation dates from the Shield database record, 1 hour intervals to extract 50 vibration data samples training the synthesized as BP network. Network input is taken as the top10 moments of the vibration values, output for the next time the vibration value. The network input layer nodes is 10 output nodes is 1, and hidden layer nodes is calculated as 10. After network converges to a given accuracy, enter any consecutive 10 moments the vibration value, the network can quickly predict the next time vibration value. Figure 5 expressed on Oct7 vibration predictive value (blue line) and its corresponding measured value (red line).

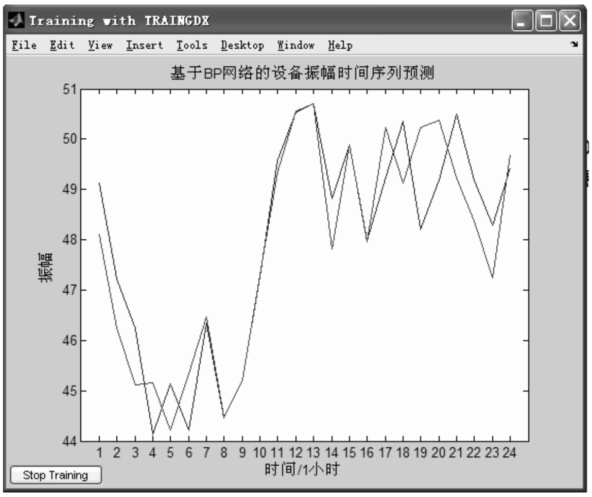

Figure 6. Synthesis of BP neural network to shield machine amplitude of time series prediction

As can be seen from Figure6, synthesis of BP network forecast with actual values is in good agreement, just with some random differences. The results show that: The synthetic prediction method of neural network can be a good future of unit operation.

7. Conclusion

This paper introduces a method which combines rough set and neural network into the fault diagnosis of Shield machine, fully playig the advantages of both. Using rough sets theory in dealing with problem, does not need to provide any prior information to deal with data sets, which can be extracted the smallest decision making rules from data, based on these rules establishing neural network with high learning rate, fewer training times, shorter training time and strong fault tolerance features. Accuracy of fault diagnosis has been improved. And using BP neural network combined with least square method to forecast fault could well reflect the unit's future operating conditions. The method used in this article has high practical value, Example shows its application to fault diagnosis of Shield machine obtains very good results.

References

[1] ZHANG Z Y. Shield Method Development and Application in China[J]. *Tunnel construction* 2006: 69-71.

[2] ZHANG F Y. *Shield tunnel[M]*. Beijing: People's Traffic Press, 2004.

[3] HUANG Z L. *Rough Set Theory and Applications [M]*. Sichuan: Chongqing University Press, 1996.

[4] ZHANG W X. *Rough set theory and method[M]*. Beijing: Science Press, 2001.

[5] Guan J W,Bell D A. Rough Computational Methods for Information System. *Artificial Intelligence*, 1998,105(1-2):77-80.

[6] GAO J. *Artificial neural network theory and simulation[M]*. Beijing: Mechanical Industry Press, 2007.

[7] FAN B B. Based on expert knowledge fault diagnosis system for shield machine[D]. *Shanghai: Shanghai University*, 2008.01.

[8] Roman W. Swiniarskia, Andrzej Skowron.Rough set methods in feature selection and recognition [J]. *Pattern Recognition Letters*. 2003 (24).

[9] MO J,WANG X, DONG M, et.Based on rough set theory for power transformer fault diagnosis[J]. *CSEE*, 24 (7) :163-164.

[10] ZHANG Y,ZHANG G L. Based on rough set and neural network fault diagnosis[J]. *Machinery and Electronics*, 29: 116-117.

[11] T. Jerome etals, Recurrent Neural Networks and Robust Time Series Prediction. *IEEE Trans. On Neural Networks*, 1994, 5(2)