Fault Diagnosis of Oil Pump Based on High Speed and Precise Genetic Algorithm Neural Network

Meijuan Gao[1,2], Jingwen Tian[1,2]
[1)] Department of Automatic Control
[1)] Beijing Union University
Beijing, China
ttgg500@163.com

Liting Cao[1], Jin Xu[2]
[2)] School of Information Science
[2)] Beijing University of Chemical Technology
Beijing, China
ttjjww999@163.com

***Abstract*—Considering the issues that the relationship between the fault of oil pump existent and fault information is a complicated and nonlinear system, and it is very difficult to found the process model to describe it. According to the physical circumstances of oil pump, a fault diagnosis method of oil pump based on high speed and precise genetic algorithm neural network is presented in this paper. The high speed and precise genetic algorithm neural network is combined the adaptive and floating-point code genetic algorithm with BP which has higher accuracy and faster convergence speed. With the ability of strong self-learning and function approach and fast convergence rate of high speed and precise genetic algorithm neural network, the diagnosis method can truly diagnose the fault of oil pump by learning the fault information of oil pump. The real diagnosis results show that this method is feasible and effective.**

***Keywords*—genetic algorithms, neural network, oil pump, fault diagnosis**

I. INTRODUCTION

The rod pumping method is a traditional mechanical oil recovery technique which is universally used domestically and abroad. At present, this method is used in about 80%~85% of the oil wells all over the world. The conventional pumping rod always do to and fro motion in the oil pipe while it is running, and this action will cause the flowing problems: wear out to rupture, low efficiency of the pump, high power cost, etc. The oil pump is an important device in petroleum industry, its working security and reliability will essentially affect the output of petroleum, the cost of power, and the economic benefit of enterprise [1]. The main faults of oil pump are: the broken of pumping rod and the eccentric wear of pumping rod. The character of a broken pumping rod is that the rod will lose its action, and can not pump out the oil, but the motor still runs, and no evident character can be seen outside. This will result in hollow pumping action which wastes much energy. The character of pumping rod eccentric wear is that the pumping rod excursion, the friction between the oil pipe and pumping rod will increase, which may cause leak, even broken of the pumping rod.

The faults of oil pump is a main problem faced by the oil pump manage department, it not only seriously effected the running of pump and the output of petroleum, but also caused large costs on the replacing of oil pipes and pumping rod broken owing to eccentric wear. Therefore, once there is fault exists in the oil pump, we must diagnose the fault as soon as possible, otherwise, it will cause the big waste of energy and will make bad influence to the production which will bring loss to enterprise. Thus, it has significant meaning to develop the fault diagnosis research of oil pump and timely discover the faults exist in the oil pump.

At present, the fault diagnosis of oil pump well is realized by manual acquisition of various measuring parameters from the locale monthly, and drawing the power curve of every oil pump well according to these parameters, then according to this power curve, manual analysis and diagnosis whether there is fault exist. This method not only cost a lot of human resource, material resource, and financial, but also can not timely diagnose the fault of oil pump. So a fault that occurred in the beginning of the month may be found until the end of the month. This seriously affects the production. Therefore, there are some other methods for diagnosis the faults exist in the oil pump were presented, such as the machine study method, especially the artificial neural network was used widely [2].

The BP network based on gradient descend is a new method in recent years, its ability to impend nonlinear function has been proved in theory also have been validated in actual application [3]. But the BP network has some problems such as converge to local minimum and the speed of network learning slowly. Genetic algorithm is also a new optimum algorithm developed fast recently, it has some advantages such as the parallel search and the searching efficiency is higher, in addition, it belongs to the random optimize process essentially, so the local convergence question is not exist. But the genetic algorithm also exist some shortage, it can but search out the approximate to excellent solution that near to global optimal solution in a short time [4].

Considering the shortage of traditional genetic algorithm and BP algorithm, a generic algorithm with adaptive and floating-point code is proposed, this algorithm is combined with BP to give high speed and precise genetic algorithm neural network which overcomes disadvantages of the genetic algorithm and BP algorithm and has higher accuracy and faster convergence speed. So in order to diagnose the faults of the oil pump effectively and in time in the petroleum production process, a fault diagnosis method based on the high speed and precise genetic algorithm neural network is presented in this paper.

II. METHOD STUDY

A. BP Network

The neurons are arranged as some layers in BP network, the network composed by one input layer and one or more hidden layers and one output layer. The learning course of network includes two courses, one is the input information transmitting forward directed and another is the error transmitting backward directed. When the information is transmitted forward directed, the input information goes to the hidden layers from input layer and goes to the output layer. If the output of output layer is different with the wishful output result then the output error will be calculated, the error will be transmitted backward directed then the weights between the neurons of every layers will be modified in order to make the error become minimum. A three layers BP network is shown in Fig 1.

The numbering of input layer is i , the numbering of hidden layer is j , the numbering of output layer is k .

Then the input of the j neuron of hidden layer is:

$$net_j = \sum_i w_{ji} o_i \tag{1}$$

The output of the j neuron is:

$$o_j = g(net_j) \tag{2}$$

The input of the neuron of output layer is:

$$net_k = \sum_j w_{kj} o_j \tag{3}$$

Corresponding output is:

$$o_k = g(net_k) \tag{4}$$

where, is Sigmoid function,

$$g(x) = \frac{1}{1+e^{-x}}$$

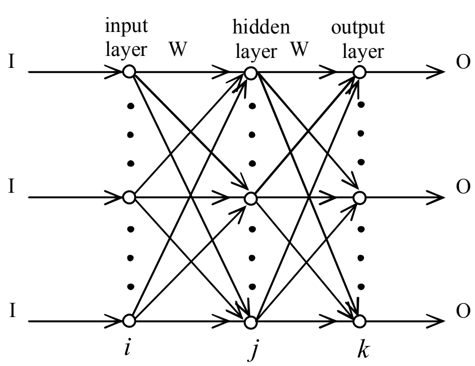

Figure 1. Structure of BP neural network

The key of BP network is the error transmitting backward directed of learning course, the course is accomplished through minimize an object function that is the error sum of squares between the actual output of network and the expectant output. Using the gradient descent algorithm we derive the computing formula.

In the learning course, supposed the expectant output of the k neuron of output layer is t_{pk} , the corresponding actual output of network is o_{pk} , then the average error of system is

$$E = \frac{1}{2p} \sum_p \sum_k (t_{pk} - o_{pk})^2 \tag{5}$$

where, p is the training samples number.

In order to the express convenience, omit the subscript p , the formula (5) becomes as follow.

$$E = \frac{1}{2} \sum_k (t_k - o_k)^2 \tag{6}$$

where, E is the object function.

According to the gradient descent algorithm, we derive the adjustment value of every weighting as follows.

$$\Delta w_{kj} = \eta (t_k - o_k) o_k (1 - o_k) o_j \tag{7}$$

$$\Delta w_{ji} = \eta \delta_j o_i \tag{8}$$

where, η is the rate of learning;

$$\delta_j = o_j (1 - o_j) \sum_k \delta_k w_{kj}$$

$$\delta_k = (t_k - o_k) o_k (1 - o_k)$$

The more bigger the rate of learning η became, the more the adjustment value of every weighting, this can accelerate the training course of network, but the result can generate oscillation, In order to avoiding the oscillation when increase the rate of learning η , adding a momentum term in the formula (7) and (8), namely:

$$\Delta w_{ji}(n+1) = \eta \delta_j o_i + \alpha \Delta w_{ji}(n) \tag{9}$$

where, α is the proportionality constant.

Through the BP network training, satisfy the accuracy requirement, then the interconnect weighing between every nodes are ascertained, here, the trained network can identify and predict the unknown sample.

B. Genetic Algorithm

The genetic algorithm (GA) is a kind of self-adapting heuristic global search algorithm which derived from imitating the thought of natural biological evolution. In nature, it is a cycle process made up of reproduction-crossover -mutation operators. In the process of searching for the global optimum solution, GA needs neither the information of gradient nor the calculus computing, it can find out the global optimum solution or near-optimal solution in the solution space with high probability only by operating the reproduction-crossover-mutation operators, thereby, it could reduce the probability of getting into the local minimum efficiently.

The reproduction operator reproduces the individuals to the new colony according to the probability in proportion as their adaptive value. After reproduction, the preponderant individuals are preserved and the inferior individuals are weed out, and the average fitness degree of the colony is increased, but the variety of colony is loss at the same time. The action of reproducing operator is to realize the principle of winner priority for preserving predominance and natural selection, and make the colony converge on the optimum solution. The crossover operator first selects two individuals stochastically according to the certain exchanging probability P_c, and then it can produce two new individuals by exchanging parts of chromogene stochastically. The genetic algorithm can generate filial generation colony which have higher average fitness and better individuals through the reproduction and crossover operators, and make the evolutionary process proceed to the optimum solution. The mutation operator changes several bits of the chromosome string stochastically with a small probability P_m, namely turn 0 to 1 and 1 to 0. The mutation operator is very important to recoup the loss of colony diversity.

C. Adaptive Genetic Algorithm (AGA)

AGA is a kind of GA that has scale reproduction and self-adaptive crossover and mutation operations. In the process of searching for the optimum parameter, AGA changes the crossover probability and mutation probability adaptively according to the different condition of individuals in order to keep the diversity of colony and prevent the premature convergence, further it can enhance the calculating speed and precision of the algorithm.

$$P_c = \begin{cases} k_1(f_{\max} - f')/(f_{\max} - f_{avg}) & \text{if } f' > f_{avg} \\ k_3 & \text{if } f' < f_{avg} \end{cases} \tag{10}$$

$$P_m = \begin{cases} k_2(f_{\max} - f)/(f_{\max} - f_{avg}) & \text{if } f > f_{avg} \\ k_4 & \text{if } f < f_{avg} \end{cases} \tag{11}$$

here, $f_{\max}$ is the biggest fitness of colony, f_{avg} is the average fitness of colony, f' is the bigger fitness of two strings used for exchange, f is the fitness of the individual to mutate.

Generally: $k_1 = k_3 = 1$, $k_2 = k_4 = 0.5$. At practical application, the value of P_c is often in range 0.5-1.0, and the P_m in range (0.005-0.05).

D. High Speed and Precise GA-BP Network

1) Encoding Mode of Chromogene

There are two main encoding mode of the genetic algorithm: binary-coding and decimal-coding. The application of binary-coding is more widely, but it has some shortages: The algorithmic precision ε lie on the length of code string and the value range of the parameter to optimize, namely $\varepsilon = (u - v)/(2^L - 1)$, here, u and v are the boundary values of the parameters to optimize, L is the length of code string. In the operation process, the code string needs to be inverted numeral system constantly, so the calculated amount is greater.

The mixed algorithm presented in this paper denote the parameters directly with decimal-coding instead of binary-coding, thus, it can avoid the encode difficulty caused by the ambiguity of the numeric area of the network. Cancelled the process of encode and decode, so enhanced the learning speed of algorithm. The importing of decimal numeric string can enhance the computational accuracy greatly under the circumstance of the length of the numeric string is invariable.

2) High Speed and Precise GA-BP Network

When the BP network converges on local minimum, the network will present serious morbidity, how to avoid the local minimum is the key problem to be solved in the application of ANN. GA is a perfect tool to optimize the BP network, which enables it avoid the local minimum and enhance the converging speed of network [5]. The generic algorithm with adaptive and floating-point code is combined with BP algorithm to optimize the BP network, it can further enhance the predicating accuracy and converging speed of BP network.

A multilayer forward BP networks were adopted in this paper, the GA regards each weights of the network as a chromosome, and the aggregation of all the weights as an individual, and a large number of individuals will be generated in the initialization phase, which is called colony. The adaptation function of GA is constructed as follow:

$$E = \frac{1}{2p}\sum_{p}\sum_{k}(t_{pk} - o_{pk})^2 \tag{12}$$

here, p is the numbers of training samples, k is the neuron numbers of the output layer, t_{pk} is the BP network output of the neuron No. k corresponding to the sample No. p, o_{pk} is the real output of the neuron No. k corresponding to the sample No. p. The adaptation function is:

$$f = \begin{cases} C_{\max} - E & E < C_{\max} \\ 0 & E \geq C_{\max} \end{cases} \tag{13}$$

here, the $C_{\max}$ can be the maximum value E of evolutionary process.

In order to ensure the stability and global convergence of the algorithm, we adopt the best reserve mechanism in the selecting operation of GA, firstly according to the roulette selecting mechanism to select, then the most fitness individual of current solution is reproduced to the next generation colony, in order to ensure the final result that obtained as soon as the GA ends is the most fitness individual of every generation appear. The steps of algorithm as follow:

Step1: Random generate N groups initial BP network weights from different space interval of real number, and regard them as the initial colony;

Step2: Preliminary train these N groups initial weights separately using BP algorithm, if there are at least one group satisfied the accuracy requirement after training, then the algorithm end, else turn step3;

Step3: Define the numeric area respectively according to the upper limit and inferior limit of the N groups network weight which have been preliminary trained, random generate $r \times N$ groups new network weights in the numeric area, these new weights in conjunction with the N groups trained weights compose an holonomic colony, there are $(r+1) \times N$ groups network weight;

Step4: Execute reproduction, crossover and mutation adaptive genetic operation on the $(r+1) \times N$ groups weights;

Step5: If there are at least one group weights can satisfy the accuracy requirement after step4, the algorithm end, else select N groups better weights from the $(r+1) \times N$ which have been exerted on the AGA, and turn step2.

The algorithm flow is shown in Fig 2.

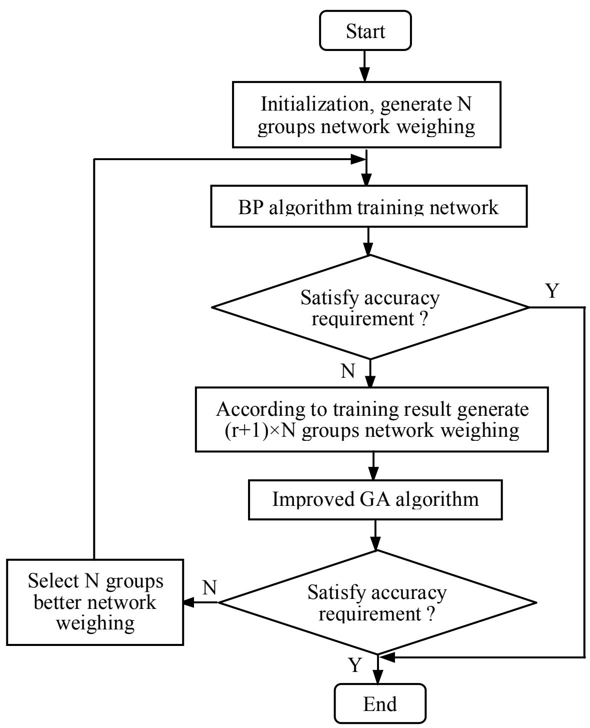

Figure 2. Flow chat of high speed and precise genetic algorithm neural network

III. APPLICATION

A. The Characteristic Parameters

The first task of fault diagnosis using high speed and precise genetic algorithm neural network is to collect sample data. In order to collect the training sample, testing sample, we must extract the characteristic parameters that correlate with fault of oil pump. In a general way, we can extract multi characteristic parameters which correlate with fault type and degree of oil pump, these characteristic parameters that we have obtained can not totally as the input vectors of sample, because among these characteristic parameters, some single characteristic parameters can not possess the direct significant, and it is possible that some characteristic parameters are correlative each other, this creates the redundancy of information, so it is necessary that optimize these characteristic parameters and select the key parameters from them. The key parameters are some minority attribute which selected from the obtained multitude characteristic parameters and have the most affiliations and the most sensitive response with the fault type and degree of oil pump. Then these characteristic parameters are analyzed and optimized. Using the forward model, we can obtain the knowledge of all kinds of characteristic parameters in reflecting the fault type and degree of oil pump, and combining the expert experience, we confirm the key parameters that have bigger degree of correlation with the fault type and degree of oil pump, the key parameters act as the input neuron of high speed and precise genetic algorithm neural network.

B. Example Research

We do the fault diagnosis of oil pump of some oil recovery plant in Daqing using the high speed and precise genetic algorithm neural network. The main faults of oil pump are the broken of pumping rod and the eccentric wear of pumping rod. In order to test the fault of oil pump in more details, we can divide the pumping rod eccentric wear into mild eccentric wear, moderate eccentric and excessive eccentric wear. Then the working state of oil pump is composed of “normal working”, “pumping rod broken”, “pumping rod mild eccentric wear”, “pumping rod moderate eccentric” and “pumping rod serious eccentric wear”.

Through the experiment, we found that the working states of the oil pump are related with the working current of oil pump, and we can estimate whether the oil pump exists the fault and the fault type by the change of the working current. Through the correlative analysis, we find the characteristic parameters which have the maximum correlativity with the fault type and degree of oil pump, combining expert experience, finally, we select 4 kind of characteristic parameters to as input neuron of high speed and precise genetic algorithm neural network, these parameters are working current, maximum current, valley current and mean square current, then the input layer of high speed and precise genetic algorithm neural network has 4 nodes. The output layer of high speed and precise genetic algorithm neural network has 5 nodes which stand for 5 kind of output mode, the expectant output of normal working state sample is [1,0,0,0,0], the expectant output of pumping rod broken state sample is [0,1,0,0,0], the expectant output of pumping rod mild eccentric wear state sample is [0,0,1,0,0], the expectant output of pumping rod moderate eccentric state sample is [0,0,0,1,0], the expectant output of pumping rod serious eccentric wear state sample is [0,0,0,0,1].

Collecting the training samples which be request to have categoricalness. We select the 80 wells signal and 20 wells signal from 100 fault wells as known training samples and diagnosis testing samples, and select the 30 wells signal and

10 wells signal from 40 normal wells as known training samples and diagnosis testing samples. We can get many group training samples and testing samples, and every training sample is composed of 4-input and 5-output. The samples data is shown in table 1.

We write the corresponding algorithm program using C++. The selection of network parameters is very flexible, however, aimed at the peculiarity of the algorithm, that is, the more the hidden layer, the more the network weights, and the more corresponding chromosome number of GA, which influence the training speed, so under the circumstance of the difference of error accuracy is less, the hidden layer of the network will not exceed 4 layers. In this paper, the structure of the high speed and precise GA-BP network is 4-12-12-5, which composed of one input layer, two hidden layer and one output layer, the input layer has 4 junction points, the every hidden layer has 12 junction points, the output layer has 5 junction points, which denote 5 kind of working state of oil pump.

In order to show the advantage and feasibility of the high speed and precise GA-BP network, we adopt the high speed and precise GA-BP network and BP network to diagnose the fault of oil pump. The structure of BP is the same as the high speed and precise GA-BP network. So the topology structure both of the high speed and precise GA-BP network and the BP network is 4-12-12-5. Training samples are inputted into the high speed and precise GA-BP network to learn, only when the BP network and the genetic algorithm both have finished the given iterative times, then the high speed and precise GA-BP network has finished once iteration. The learn parameters are defined as follows: the learn speed is 0.7, inertia coefficient is 0.8, the maximum error of system is 0.01, the maximum error of single sample is 0.001, BP iterative times are 30, the improved genetic algorithm iterative times is 20, the high speed and precise GA-BP network iterative times are 5000. When the GA-BP network has been trained to the required target, the network is confirmed.

In addition, we select different training sample numbers and different test samples numbers every time and input the training samples into the high speed and precise GA-BP network and BP network to learn, then use the test samples to test the result. The result of train and test by the high speed and precise GA-BP network and BP network is shown in table 2.

TABLE I. SAMPLE DATA

Fault type \ Input node	x_1	x_2	x_3	x_4
Normal	65.3869	55.2312	53.5232	57.5046
pumping rod broken	50.6523	49.2135	44.8231	42.9532
pumping rod mild eccentric wear	61.5006	57.2315	54.2135	52.3675
pumping rod moderate eccentric wear	61.0432	59.6572	55.4323	52.9392
pumping rod serious eccentric wear	60.6521	60.5782	57.6783	53.6678

TABLE II. MEAN SQUARED ERRORS ANALYSIS OF GA-BP AND BP

train sample	test sample	high speed and precise GA-BP network		BP network	
		mean squared errors of train sample	mean squared errors of testing sample	mean squared errors of train sample	mean squared errors of testing sample
60	20	0.079	0.112	0.096	0.196
80	30	0.071	0.126	0.089	0.227
100	40	0.060	0.143	0.079	0.305
120	50	0.053	0.169	0.072	0.352

From the table 2, we can see the mean squared errors of training samples of the high speed and precise GA-BP network is smaller than that of BP network. Moreover the mean squared errors of testing samples of the high speed and precise GA-BP network is also smaller than that of BP network. When the number of the training samples change, we can get the same result. It shows that the fault diagnosis method based on the high speed and precise GA-BP network has higher stability, and can obtain higher diagnosis accuracy and faster diagnosis speed.

IV. CONCLUSION

The high speed and precise genetic algorithm neural network is applied to fault diagnosis of oil pump, it combines the advantages of both GA and BP network, and it can overcome the disadvantages of traditional BP algorithm such as local minimum and the low converging speed, therefore, the converging speed and the diagnosis accuracy of system are enhanced. With the ability of strong self-learning and function approach and fast convergence rate of high speed and precise genetic algorithm neural network, the diagnosis method can truly diagnose the fault of oil pump by learning the fault information of oil pump. This method not only can save a lot of human resources, material resource, and financial, but also can timely diagnose the fault of oil pump and make the possible loss decrease to least. The real fault diagnosis results show that this method is feasible and effective and this method has great application foreground in the oil production.

ACKNOWLEDGMENT

This work was supported in part by the National High Technology Research and Development Program of China (863 Program) under Grant #2005AA420070-1 and Beijing Education Special Project under Grant PXM2008_014209_054622.

REFERENCES

[1] Wu Yitong, Liu Zhuojun, Zhao guixiang, et al, “Oil pump,” Beijing: petroleum industry publishing house, 1994.

[2] Liu Mushuang, Cai Guangxin, Xie Ying, “Long distance intelligent fault diagnosis for walk-beam pumping units,” Oil Field Equipment, vol. 35, no. 5, pp. 15-18, 2006.

[3] Jiao Licheng, “Neural network system theory,” Xian: Xi an electronic science and technology university press, 1995.

[4] Zhao Zhenning, Xu Yongmao, “Introduction to fuzzy theory and neural networks and their application,” Beijing: Tsinghua University Press, 1996.

[5] Grefenstetle J., “Optimization of control parameters for genetic algorithm,” IEEE Trans on Syst, Man and Cybern, vol. 16, no. 1, pp. 122-128, 1986.

[6] Huang Xiuxuan, Zhang Xuefeng, “Modified adaptive genetic algorithm,” Academic Periodical Abstracts of China, vol. 14, no.11, pp. 1415-1417, 1998.

[7] Wang Jiancheng, Gao Daqi, et al, “Designing of ANN economic early warning system based on improved genetic and BP algorithm,” Systems Engineering-Theroy & Practice, vol. 18, no. 4, pp. 136-141, 1998.

[8] Bian Zhaoqi, Zhang Xuegong, “Pattern recognition,” Beijing: Tsinghua university publishing house, 2000.

[9] Li Shiyong, “Fuzzy Control Neurocontrol and Intelligent Cybernetics,” Harbin: Harbin Institute of Technology publishing house, 1996.

[10] Steyer Jean Philippe, Rolland Damien, Bouvier Jean Claude, Moletta René, “Hybrid fuzzy neural network for diagnosis application to the anaerobic treatment of wine distillery wastewater in a fluidized bed reactor,” Water Science and Technology, vol. 36, no. 627, pp. 209-217, 1997.

[11] Jiang Jianguo, et al, “fualt diagnosis and application in electrician,” Beijing: science publishing house, 1995.

[12] Wu Guifeng, Zhai Yuqing, Chen Hong, et al, “Fault Diagnosis of Vibration in Electical Machine Based on the Wavelet neural Network,” Control Engineering of China, vol. 11, no. 2, pp. 152-154, 2004.