Fault Diagnosis for Vehicle on-Board Equipments of High-Speed Railway Based on GA-ANFIS Model

Linggang KONG [a,1], Shuo LI [a], Xinlong CHEN [a] and Hongyan QIN [b]

[a] *National Research Center of Green Coating Technology and Equipment Engineering Technology, Lanzhou*
[b] *Harbin West Railway Station Headquarters, Harbin Railway Bureau, Harbin 150000, China*

Abstract. Vehicle on-board equipment is the most important train control equipment in high-speed railways. Due to the low efficiency and accuracy of manual detection, in this paper, we propose an intellectualized fault diagnosis method based on adaptive neuro-fuzzy inference system (ANFIS) network. Firstly, we collect the fault information sheets that are recorded by electrical personnel, using frequency weighting factor and principal component analysis (PCA) to realize the data extraction and dimension reduction; Then, in order to improve the fault diagnosis rate of the model, using genetic algorithm (GA) to optimize the parameters of the ANFIS network; Finally, using the fault data of a high-speed railway line in 2019 to test the model, the optimized ANFIS model can achieve 96% fault diagnosis rate for vehicle on-board equipments, which indicating the method is effective and accurate.

Keywords. Vehicle on-board equipments, fault diagnosis, genetic algorithm, ANFIS

1. Introduction

In China, the railway has an important tool of transportation; according to the statistics of China Railway Corporation, the operating mileage of China's railway has reached 139000 km, of which the high-speed railway is 35000 km, ranking first in the world. China train control system is one of the key technologies to ensure the safe operation of high-speed railway. There are three types of VOBEs used in China, CTCS-300T, CTCS-300H and CTCS-300S; among them, CTCS-300T is the most widely used in the China. The paper takes this as the research object.

When the train moves in high-speed, VOBEs happened failure and can not be solved in time, it will cause serious consequences. Because of the complex structure of VOBEs, which can not meet the requirements of HSRs only by manual detection and maintenance. So, it is necessary to study the fault diagnosis technology of VOBEs with high reliability and high performance.

---

[1] Corresponding Author：Kong Linggang,（1978-）Male,AnHui,Associate Professor E-mail: konglinggang1978@163.com

After decades of development of intelligent analysis and fault diagnosis technology, its theoretical basis has been involved in artificial intelligence, automatic control, data mining and other disciplines. The intelligent analysis and fault diagnosis models are as follows: the case-based intelligent analysis and fault diagnosis models are as follows: The inference [2] proposed the Internet big data information analysis and power intelligent risk prediction based on case-based reasoning. The main idea is to combine the data mining technology of NT-SMOTE with the case-based reasoning technology, and constructing a case-based reasoning based on closed triangle. The intelligent analysis and fault diagnosis models based on the analytical model are as follows: In reference [3-4], an optimization model for fault diagnosis based on the physical model of power grid information is proposed, and the problem is expressed as an integer programming problem whose objective function is the extreme value; The analysis and fault diagnosis models based on artificial intelligence are as follows: the main idea of literature [5-7] is intelligent diagnosis based on neural network, and the fault phenomenon is used to mine the fault causes.

Based on the idea of intelligent fault diagnosis, the complexity of train control equipment structure and the diversity of fault types, paper proposes a fault diagnosis method based on adaptive fuzzy neural network. The basic idea is as follows: Firstly, We will process the fault information tables which was collected from electricians and converted into doc characteristics [8], used principal component analysis (PCA) to realize the data extraction and dimension reduction Then, in order to improve the generalization ability of the model, using genetic algorithm (GA) to optimize the parameters of the ANFIS network so that the model can converge better, Finally, using the fault data of a high-speed railway line to test the model.

2. Problem Statement

2.1. The Structure of VOBEs

The paper focuses on the research of CTCS-3 300T VOBE, which adopts the distributed structure, and each module is connected by MVB bus and Profi bus. The VOBE mainly includes track circuit information receiving unit (TCR), vehicle safety computer unit (VCU), Driver-Machine Interface (DMI), train signal gateway (TSG), balise transmission module (BTM),Global System for Mobile Communications-Railway (GSM-R), digital input / digital output (DI/DX).

2.2. The Structure of ANFIS

Adaptive neuro-fuzzy inference network (ANFIS) which is a fuzzy reasoning system based on Takagi Sugeno model. It combines neural network with fuzzy logic, and can automatically generate if-then rules [10]. A typical ANFIS structure is shown in Figure 1.
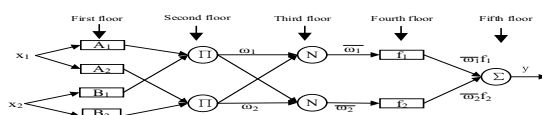

Figure 1. Structure diagram of ANFIS

We will assume two inputs x and single output y as examples in the paper to expound the principle of ANFIS structure chart. Suppose that the rule base contains two fuzzy if-then rules of Takagi Sugeno type.

Rule 1 :   If x is A$_1$ and y is B$_1$,   then $f_1 = p_1 x + q_1 y + r_1$,

Rule 2 :   If x is A$_2$ and y is B$_2$,   then $f_2 = p_2 x + q_2 y + r_2$ 。

Where p_i、q_i、r_i is the conclusion parameter of rule 1 and p_2、q_2、r_2 is the conclusion parameter of rule 2.The structure of ANFIS will not be described here. Please refer to [10] for details.

3. Modeling Based on GA-ANFIS

3.1. *Optimization of ANFIS Parameters by Genetic Algorithm*

At present, common hybrid learning algorithm is used to train ANFIS network parameters. Compared with other network models, it will have better convergence speed. However, due to the back-propagation algorithm, the convergence speed will be affected, so GA is needed to optimize its parameters. The fuzzy neural network parameters include: the center value of Gauss membership function, the width and the consequent parameters. The main idea of GA in [11] is to select individuals by fitness function, and carry out crossover and mutation operations to keep the individuals with good fitness and eliminate the individuals with poor fitness, so that the new generation of individuals not only retains the genes of the previous generation, but also is superior to the previous generation. This is repeated until the conditions are met. The specific steps are as follows:

   ● Generate an initial population randomly. And each individual is a real number string.

   ● The MSE function of adaptive fuzzy neural network is used as the fitness function of genetic algorithm.

$$F(x_i) = MSE = \frac{1}{n}\sum_{i=1}^{n}|f(x_i) - y(x_i)|$$

(1)

Where, $F(x_i)$ is the fitness function, $f(x_i)$ is the expected output, $y(x_i)$ is the predicted output, n is the output node.

   ● Population selection, crossover, and mutation operations. The selection operation uses roulette method. The higher the fitness, the greater the probability of selection. The probability of selection is expressed by:

$$P(x_i) = \frac{F(x_i)}{\sum_{i=1}^{n}F(x_i)}$$

(2)

Randomly select two individuals a_k and a_l from the previous generation to perform crossover operation at position j :

$$a_{kj} = a_{kj}(1-\beta) + a_{kj}\beta$$
(3)

$$a_{lj} = a_{lj}(1-\beta) + a_{lj}\beta$$
(4)

Among them, β is a random number between (0~1). Select the i gene of the j individual for mutation operation:

$$a_{ij} = a_{ij} + (a_{ij} - a_{min})\gamma \qquad \gamma \le 0.5$$
(5)

$$a_{ij} = a_{ij} + \left(a_{ij} - a_{max}\right)\gamma \qquad \gamma > 0.5$$
(6)

Optimal preservation strategy: When a new generation of population individuals is finally obtained, if the minimum fitness value of an individual is less than the maximum fitness value of the previous generation, the maximum fitness value of the previous generation will replace the minimum fitness value of the new generation, Can avoid the result falling into the local optimal solution. The optimal individual obtained is used as the parameter of the adaptive fuzzy neural network.

3.2. On-board Fault Data Preprocessing

The fault data of train control on-board equipment is mainly the fault information table recorded by the electric service personnel and the AElog file recorded by the on-board computer. Because the current vehicle fault data is described by text, lack of unified terminology and concepts, resulting in the lack of standardization of fault description, so it is necessary to preprocess the fault data and extract characteristic vocabulary [11]. The article analyzes the failure information table of Harbin Railway Bureau in 2019. Statistics are made on the modules with higher frequency of failure. As shown in Figure 2.
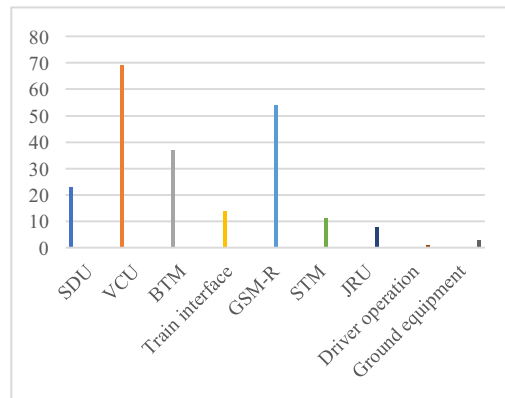

Figure 2. Statistical chart of failure times

It can be seen from Figure 2 that there are many failures in SDU, VCU, GSM-R, BTM, and train interfaces. The failure type table is shown in Table 1 below. A large

number of characteristic vocabularies can be obtained according to the fault information table and the proper nouns of the train control on-board equipment, and then the expert knowledge and related data are used to filter and obtain the following characteristic vocabulary representing the fault symptoms. The fault characteristic vocabulary is shown in Table 2.

Table 1. Fault types

Number	Fault style	Number	Fault style
A	SDU	B	VCU
C	GSM-R	D	BTM
E	Train interfaces		

Table 2. Glossary of fault characteristics

Number	Characteristic vocabulary	Number	Characteristic vocabulary
1	ATP failure	2	C3 to C2
3	Park	4	BTM failure
5	Failed to start	6	Security software failure
7	Balise message error	8	Interruption between host and DMI
9	All zero message	10	Wireless connection timeout
11	SDU failure	12	Switch system restart

The fault information table can be expressed as a word set. As word frequency can not fully explain the importance in the text. Therefore, TF-IDF method is adopted to calculate the weight of words in the document. The method not only considers the weight of a word in a single text, but also thinks over the frequency of the word appearing in the whole text [12]; as shown in formula (7), calculate the weight t_i in document d_i

$$w_{ik} = \frac{tf_{ik} \times \log\left(\frac{N}{n_k} + 0.1\right)}{\sqrt{\sum_{i=1}^{n} tf_{ik} \times \log\left(\frac{N}{n_k} + 0.1\right)^2}} \tag{7}$$

According to the above method, the weight of each feature word under each entry is calculated. If the feature word is not included in the entry, the weight is 0. According to table 1 and table 2, the weight table of fault feature words in Table 3 is obtained by frequency weighting factor method.

Table 3. Weight table of fault text feature attribute

Number	Characteristic words						Fault Type
	1	2	3	4	5	6	
Text 1	0. 25	0. 11	0. 078	0	0	0	B
Text 2	0	0	0. 045	0	0	0	E
Text 3	0. 178	0	0	0. 036	0	0	D
Text 4	0	0. 57	0	0	0	0	C
......							
Text 220		0.0146			0.369		E

For the fault text weight table, there are a lot of data redundancy characteristics, which will make the network structure more complex. Therefore, the introduction of principal component analysis (PCA) which aims to eliminate redundant attributes without losing any useful information, so as to achieve the reduction effect, and replace the original variables with some new variables, The new variable is required to reflect the information of the original variable as much as possible, extract the main components of the data, and to make the main feature components maintain the integrity of the whole data information. The results are shown in Figure 4.
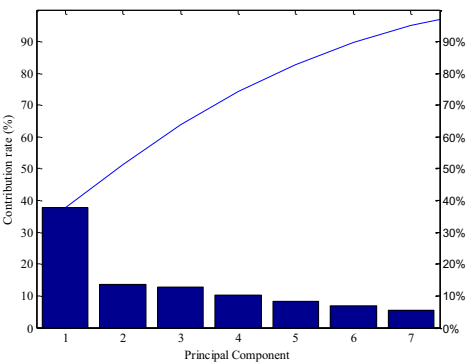

Figure 4. Results of principal component analysis

As can be seen from Figure 5, the contribution rate of the first seven principal components reaches 96.38% > 85%, which can approximately replace other components, so as to achieve the purpose of dimensionality reduction. The result of reduction is {ATP fault, interruption of communication between parking host and DMI, all zero message, balise message error, BTM fault, SDU fault, system switching restart}. Through principal component analysis, the original 12 fault feature words are reduced to 7, which are used as the input of neural network. Compared with the previous 12 inputs, the purpose of simplifying the model is achieved. The five feature words that are eliminated are approximately regarded as useless information, which will not have a great impact on the results.

3.3. Network Construction

Matlab is used as the simulation software to construct GA-ANFIS neural network. In this paper, the ANFIS parameters are set as follows: the membership function of input variables is Gauss function, the number of membership functions is 3, the training and learning algorithm uses BP algorithm, the number of iterations is set to 500 times, the error is 1×10^{-4} GA parameter setting: the initial population number is 50, the crossover probability is 0.8, and the mutation probability is 0.001.

The construction of the whole network is divided into two stages of training and testing. 180 groups of data are randomly selected to build the network and load the training data samples. The output model is selected as follows: 1 is SDU related fault, 2 is VCU related fault, 3 is GSM-R related fault, 4 is BTM related fault, 5 is train interface related fault. Then, the training samples are used to train the model for many times. Then the remaining 40 groups of data are used to verify whether the fault diagnosis rate and convergence condition of the model meet the requirements.

4. Experimental Results and Analysis

As shown in Figure 5 and Figure 6, the simulation results of training phase and test phase are shown respectively. It can be seen from the figure that the average fault recognition rate of ANFIS and GA-ANFIS in training stage is 31.91% and 96.369%, that of ANFIS and GA-ANFIS in testing stage is 23.229% and 95.46%. The rate is better than the training stage, which shows that the sample division is more accurate and the model is well constructed. The optimized ANFIS fault identification rate is increased by three times, and the optimized model is more stable, and the curve fluctuation is smaller. It shows that genetic algorithm optimizes the parameters of adaptive fuzzy neural network, which makes ANFIS convergence faster and error smaller.
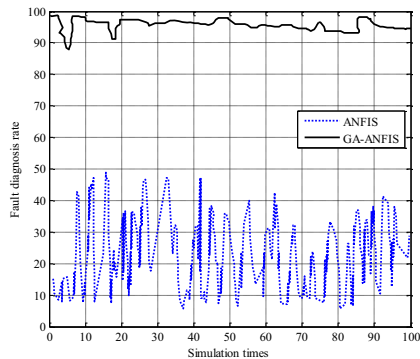

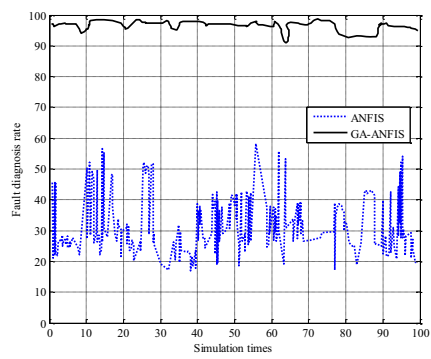**Figure 5.** Result chart of training stage   **Figure 6.** Result chart of testing stage

Table 4. GA-ANFIS output value

Ideal Output Value	Actual Output Value	Average Error
1	0. 9857	
2	1. 9895	
3	3. 0136	0.11%
4	4. 0038	
5	4. 9936	

Table 5. ANFIS output value

Ideal Output Value	Actual Output Value	Average Error
1	0.8487	
2	2.2784	
3	3.0249	2.81%
4	3.8473	
5	4.8729	

Table 4 and 5 are the test output values of GA-ANFIS and ANFIS. It can be seen from Table 4 that the error between the ideal output and the actual output of the

optimized neural network is small. In summary, it can be seen that GA-ANFIS has successfully identified five kinds of faults of train control on-board equipment.

5. Conclusion

● The combination of genetic algorithm and adaptive fuzzy neural network can greatly improve the convergence speed of ANFIS network.

● When dealing with fault text data, frequency weighting factor and expert knowledge are used to get the weight table of feature words, and then principal component analysis is used to reduce the dimension of feature words to get a simplified model.

● The experimental results show that the fault diagnosis rate of optimized ANFIS can reach 96.396%, which is 3 ~ 4 times higher than that of the non optimized ANFIS model. It is proved that the GA-ANFIS model proposed in this paper can accurately identify the fault of train control on-board equipment, and can provide reference for the intelligent diagnosis method of post high-speed railway.

References

[1]   Dong Y. Interval signal and train operation control system. Beijing: China Railway Press, 2008, 180-213.

[2]   Xu YH, Tian XY. Internet big data information analysis and power intelligent automation risk prediction based on case based reasoning. Proceedings of the 2015 3rd International Conference on Machinery, Materials and Information Technology Applications.

[3]   Zhang XY. Research on intelligent vehicle fault diagnosis system based on case similarity. Jilin University.

[4]   Dai X. Optimization model of fault diagnosis based on cyber physical model. Journal of Electric Power, 2019, 4(02): 158-166.

[5]   Shang Guan EE, Meng YY, Yang JM, Cai BG, LSTM-BP neural network based fault diagnosis for on-board equipment of Chinese train control system. Journal of Beijing Jiaotong University, 2019, 43(01): 54-62.

[6]   Liang X, Wang HF, Guo J, et al. Bayesian network based fault diagnosis method for on-board equipment of. train control system. Journal of China Railway Society, 2017, 39(08): 93-100.

[7]   Ding S, Xu X, Zhu H, et al. Studies on Optimization Algorithms for Some Artificial Neural Networks Based on Genetic Algorithm (GA) train control system[J]. Journal of China Railway Society, 2017, 39(08): 93-100.

[8]   Zhou LJ, Dong Y. Research on fault diagnosis method for n-board equipment of train control system based on GA-BP neural network. Journal of Computers, 2011, 6(5): 939-946.

[9]   Zhang SG. CTCS-3 train control system overall technical scheme. Beijing: China Railway Press, 2008: 198-218.

[10]   Xu AD, Fan YH, Li ZQ. Modeling of switched reluctance motor based on GA-ANFIS. Journal of Electrical Engineering and Control, 2011, 015(007): 54-59.

[11]   China Railway Corporation.Typical fault case of train control on-board equipment. Beijing: China Railway Press, 2013: 13-102.

[12]   Hua WY, Zai SN, Fu ZZ, et al. Robust Recommendation Algorithm Based on Kernel Principal Component Analysis and Fuzzy C-means Clustering, 2018, 023(002): 111-119.