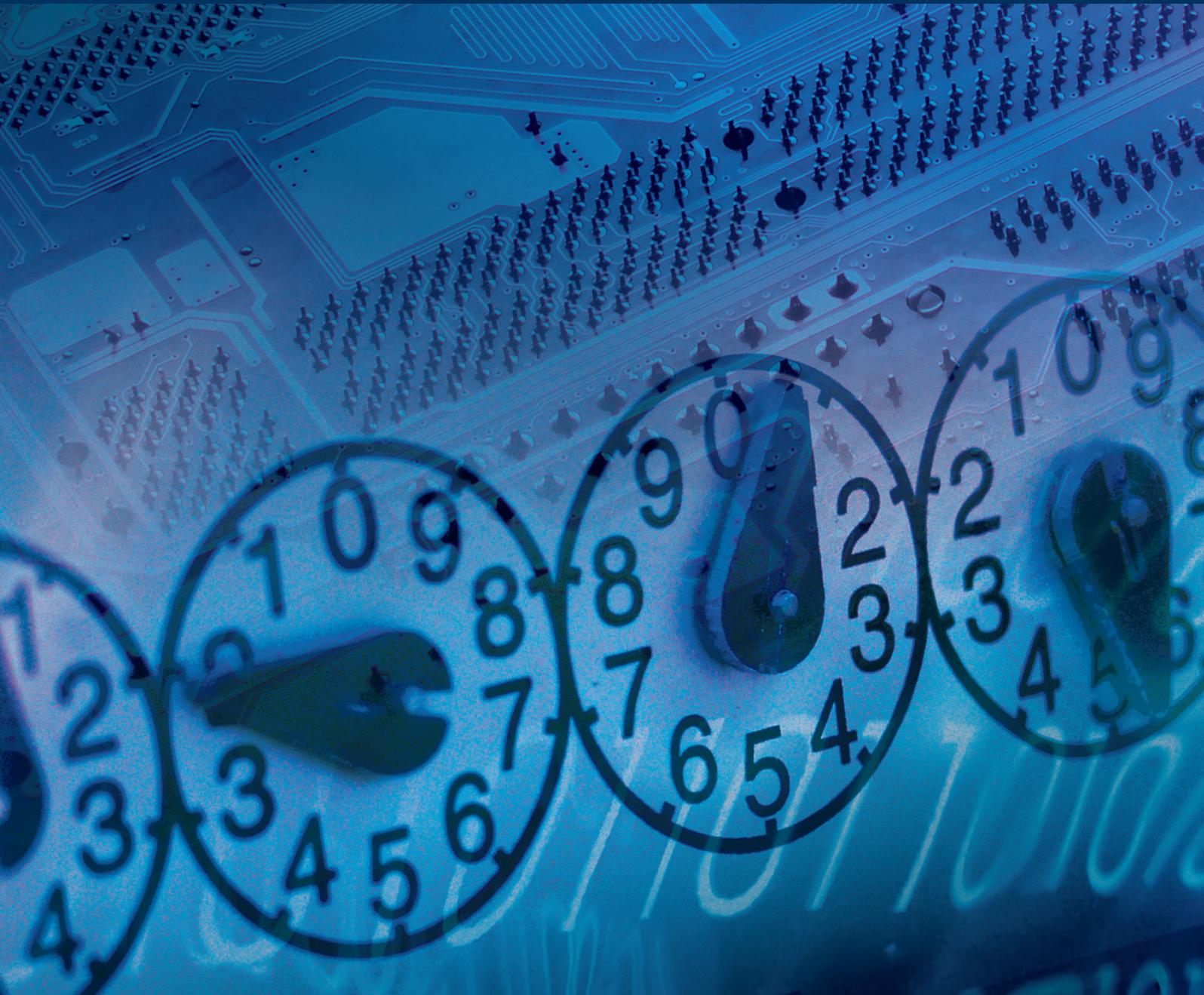


Fault Diagnosis and Application to Modern Systems 2018

Lead Guest Editor: Xiao He

Guest Editors: Zidong Wang, Zhijie Zhou, Youqing Wang,
and Qiang Liu





**Fault Diagnosis and Application to
Modern Systems 2018**

Journal of Control Science and Engineering

**Fault Diagnosis and Application to
Modern Systems 2018**

Lead Guest Editor: Xiao He

Guest Editors: Zidong Wang, Zhijie Zhou, Youqing Wang,
and Qiang Liu



Copyright © 2018 Hindawi. All rights reserved.

This is a special issue published in “Journal of Control Science and Engineering.” All articles are open access articles distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Editorial Board

Raquel Caballero-Águila, Spain
Hung-Yuan Chung, Taiwan
Ming-Cong Deng, Japan
Haiping Du, Australia
Ricardo Dunia, USA
Meng Joo Er, Singapore
Peilin Fu, USA
Furong Gao, Hong Kong
Carlos-Andrés García, Spain
Fengshou Gu, UK
Benoit Jung, France
Seiichiro Katsura, Japan
James Lam, Hong Kong
William MacKunis, USA
Radek Matušů, Czech Republic

Shengwei Mei, China
Juan-Albino Méndez-Pérez, Spain
Paolo Mercorelli, Germany
José Sánchez Moreno, Spain
Daniel Morinigo-Sotelo, Spain
Sing Kiong Nguang, New Zealand
Enrique Onieva, Spain
Belkacem Ould Bouamama, France
Petko Petkov, Bulgaria
Manuel Pineda-Sanchez, Spain
Daniela Proto, Italy
Ruben Puche-Panadero, Spain
Gerasimos Rigatos, Greece
Michela Robba, Italy
Mario Russo, Italy

Roberto Sacile, Italy
Yang Shi, Canada
Pierluigi Siano, Italy
Gangbing Song, USA
Zoltan Szabo, Hungary
Javad M. Velni, USA
Kalyana C. Veluvolu, Republic of Korea
Antonio Visioli, Italy
Yongji Wang, China
Ai-Guo Wu, China
Wen Yu, Mexico
Gang Zheng, France
Mohamed Zribi, Kuwait

Contents

Fault Diagnosis and Application to Modern Systems 2018

Xiao He , Zidong Wang , Zhijie Zhou , Youqing Wang , and Qiang Liu 
Editorial (3 pages), Article ID 8460541, Volume 2018 (2018)

Deep Sparse Autoencoder for Feature Extraction and Diagnosis of Locomotive Adhesion Status

Changfan Zhang , Xiang Cheng, Jianhua Liu, Jing He , and Guangwei Liu
Research Article (9 pages), Article ID 8676387, Volume 2018 (2018)

Data Preprocessing Method and Fault Diagnosis Based on Evaluation Function of Information Contribution Degree

Siyu Ji  and Chenglin Wen 
Research Article (10 pages), Article ID 6565737, Volume 2018 (2018)

Fault Diagnosis Method Based on Gap Metric Data Preprocessing and Principal Component Analysis

Zihan Wang, Chenglin Wen , Xiaoming Xu , and Siyu Ji 
Research Article (9 pages), Article ID 1025353, Volume 2018 (2018)

A New Residual Life Prediction Method for Complex Systems Based on Wiener Process and Evidential Reasoning

Xin Zhang, Shuaiwen Tang, Taoyuan Liu, and Bangcheng Zhang 
Research Article (12 pages), Article ID 9473526, Volume 2018 (2018)

A New Synergistic Forecasting Method for Short-Term Traffic Flow with Event-Triggered Strong Fluctuation

Darong Huang , Zhenping Deng , and Bo Mi
Research Article (8 pages), Article ID 4570493, Volume 2018 (2018)

Detection of Two-Level Inverter Open-Circuit Fault Using a Combined DWT-NN Approach

Bilal Djamal Eddine Cherif  and Azeddine Bendiabdellah
Research Article (11 pages), Article ID 1976836, Volume 2018 (2018)

Neural Back-Stepping Control of Hypersonic Flight Vehicle with Actuator Fault

Qi Wu  and Yuyan Guo 
Research Article (5 pages), Article ID 2198423, Volume 2018 (2018)

Editorial

Fault Diagnosis and Application to Modern Systems 2018

Xiao He ,¹ Zidong Wang ,² Zhijie Zhou ,³ Youqing Wang ,⁴ and Qiang Liu ⁵

¹Department of Automation, Beijing National Research Center for Information Science and Technology (BNRist), Tsinghua University, Beijing 100084, China

²Department of Computer Science, Brunel University London, London UB8 3PH, UK

³Xi'an High-Tech Institute, Xi'an 710025, China

⁴College of Electrical Engineering and Automation, Shandong University of Science and Technology, Qingdao 266590, China

⁵State Key Laboratory of Synthetical Automation for Process Industries, Northeastern University, Shenyang, Liaoning 110819, China

Correspondence should be addressed to Xiao He; hexiao@tsinghua.edu.cn

Received 2 September 2018; Accepted 2 September 2018; Published 2 December 2018

Copyright © 2018 Xiao He et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Fault diagnosis (FD) is a powerful means to guarantee the continuously increasing requirements from the modern industrial systems on better performance, more reliability, and higher safety. Therefore, the field of research on FD has been keeping vitality for the past several decades. FD utilized the analytical redundancy relationship of the control system to provide the whole system with useful and accurate information about the faults, to make further decisions such as recovering or compensating the influence of faults, predicting the residual life and so forth. A residual signal needs to be constructed and then to be compared with a preset limit for detecting the occurrence of fault. The exact location of the fault should be ensured to avoid the propagation of the fault to other components. The amplitude and emerging time of the fault also need to be estimated. All aforementioned procedures are the three basic units involved in a typical FD process, which are fault detection, fault isolation, and fault estimation. The precise investigation and comprehension of the fault achieved by the FD techniques bring a significant impact on the further process of fault accommodation.

The current available FD methodologies can be generally categorized as the model-based method and the data-driven approach. With a mathematical model, the model-based method can furnish the factors beyond the objective faults with the precise decoupling or maximum attenuation. In the case that the mathematical model is not available, however, the historical data are sufficient, and the data-driven approach reveals its superiority. The theoretical challenges as well as practical applications brought by the modern

sophisticated real-time control systems propel the research on FD. Consequently, the abundant results for the topics of FD have recently come up. To this end, we have organized a special issue Fault Diagnosis and Application to Modern Systems in 2017. Based on the success of the 2017's special issue, the editorial office has made this special issue into one of the annual special issues, which simply means it becomes the first issue in a series of special issues which will be published each year. In this special issue of the year 2018, we have received 19 manuscripts from all over the world. With careful and rigorous peer-review, seven papers have been selected and accepted to be finally published in this special issue. These seven papers included in this special issue which reflect the development in FD of modern systems will be overviewed in the following.

Some work in this special issue addresses challenges experienced in the data-driven FD techniques. In "Fault Diagnosis Method Based on Gap Metric Data Preprocessing and Principal Component Analysis", Z. Wang et al. study the use of Gap metric as a method to improve the data preprocessing in the Principal Component Analysis (PCA) approach and propose a novel PCA method for fault diagnosis. The Gap metric is capable of measuring the distance between two linear systems and reflecting the correlation of different variables of the system in high-dimensional Riemannian space. With the adoption of the Gap metric data preprocessing method, the variable data in the system can be emphasized with small absolute change and fairly large variation. Therefore, the principal component can be accurately extracted for further calculating the T^2 and SPE

statistical limits so as to detect the occurrence of fault. The results of the numerical simulation show that the proposed PCA method based on the Gap metric has better performance in omissive judgement of fault, compared with the traditional PCA method and the one in European space. Furthermore, the proposed solution is feasible even on detection of micro faults.

Another paper “Data Preprocessing Method and Fault Diagnosis Based on Evaluation Function of Information Contribution Degree” by S. Ji and C. Wen introduces the evaluation function to adaptively update the network parameters and proposes a novel method for evaluating the parameters of the function. Their proposed method uses Kalman filter to real time update the hidden layer output weights of BP neural network and adjust the output of BP neural network hidden layer, which increases the accuracy of the obtained evaluation function and avoids the waste of resources caused by the unnecessary retraining of the network. This paper further employs the Mean Impact Value (MIV) algorithm to calculate the weight of the feature, whose main purpose is to bring the weighted data into the established fault diagnosis model for FD. Their proposed FD method is further utilized in the data collected by a Combined Cyber Power Plant over 6 years. The experimental results demonstrate that the proposed method can effectively detect the faults by improving the performance of algorithm and reducing the wasteful time-consuming of retraining.

Some work focuses on the model-based FD techniques and considers the system recovery and further prediction. In the paper entitled “Detection of Two-Level Inverter Open-Circuit Fault Using a Combined DWT-NN Approach” by B. D. E. Cherif and A. Bendiabdellah, a novel diagnostic method is proposed for detecting an inverter IGBT open-circuit switch fault, based on the discrete wavelet transforms (DWT) algorithm and the approach of neural network (NN). The Coiflet algorithm is applied in the DWT, and then the stator current signal is decomposed into approximations and details that can respectively represent the slow and fast variations of the signal. In order to realize the automatic fault diagnosis, this paper further introduces the NN approach whose construction is established via the Lunemberg algorithm. The experimental results of a built voltage inverter fed induction motor illustrate the feasibility and effectiveness of the proposed diagnosis method.

The accurate residual life prediction at the beginning of performance degradation can prevent the complex systems from suffering the serious fault or major damage emerging. The paper “A New Residual Life Prediction Method for Complex Systems Based on Wiener Process and Evidential Reasoning” by X. Zhang et al. presents a novel residual life prediction method for complex systems through using Wiener process and evidential reasoning. Based on Wiener process, a degradation model is built whose parameters are estimated and updated by maximum likelihood method and Bayesian method, separately. The evidential reasoning rule is applied to establish the residual life prediction model and is further integrated with the prediction results of Wiener process to obtain more accurate prediction results. The results of the case study of the residual life of gyroscope demonstrate

that the proposed prediction method is more feasible and effective compared with the method based on fuzzy theory.

The demands for the reliability of short-term traffic flow prediction grow up with the pervasive coordination of various kinds of traffic equipment. Unfortunately, the serious fluctuation in the traffic data can deteriorate the prediction accuracy. The paper “A New Synergistic Forecasting Method for Short-term Traffic Flow with Event-triggered Strong Fluctuation” by D. Huang et al. proposes a novel short-term traffic forecasting method to solve the aforementioned problem. The traditional grey Verhulst model is improved by introducing the Markov chain. Then the authors employ the first order difference exponential smoothing technique to improve the prediction smoothness with accuracy. Finally the improved grey Verhulst prediction method and the first order difference exponential smoothing technique are synthesized by the dynamic weighting factors to achieve the goal of accurately and effectively forecasting short-term traffic flow. The proposed method is verified through a case study of the urban expressway under a strong fluctuation environment, where the comparative analysis indicated its feasibility.

Some work pays attention to the application of FD and Fault-Tolerant Control (FTC) techniques for some specific real-time systems. In “Deep Sparse Auto Encoder for Feature Extraction and Diagnosis of Locomotive Adhesion Status” by C. Zhang et al., a sparse auto-encoder deep neural with dropout is presented to tackle the FD problem of the wheel-rail adhesion state of a locomotive. Adhesion is a phenomenon caused by the rolling wheel pair subjected to a tangential traction, where the rolling pressure leads to deformation between the wheel and the rail as well as the continuous and stable contact surface between the wheel and the rail. This paper refines the adhesion features as four categories for wheel skid warning. This paper further applies an unsupervised learning algorithm that is the sparse auto-encoder and stacks it into a deep structure, which constructs a model for effectively extracting data features and making classification easier. Then, the adhesion state of the locomotive can be diagnosed based on the inherent characteristics of the sample data of the sparse auto-encoder deep neural network. The experimental results indicate that the proposed method can achieve a 99.3% diagnosis accuracy and simultaneously meet the actual engineering monitoring requirements.

The paper entitled “Neural Back-stepping Control of Hypersonic Flight Vehicle with Actuator Fault” by Q. Wu and Y. Guo addresses the FTC problem of hypersonic flight vehicle, based on a back-stepping control law with neural networks and adaptive method. The neural networks approach is utilized in the controller design so as to estimate the unknown function in flight dynamics. Considering the aspect of eliminating the influence of the actuator dead-zone fault, this paper introduces an adaptive signal in the controller design to estimate the unknown fault parameters. Moreover, the stability analysis and simulation are provided to indicate the good performance of proposed adaptive fault tolerant method.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this article.

Acknowledgments

This editorial was partially supported by the National Key Research and Development Program under Grant 2017Y-FA0700300, the NSFC under Grants 61473163, 61522309, and 61733009, and the R&D Project of Intelligent Ship 1.0 from the Ministry of Industry and Information Technology of China under Grant [2016]544.

*Xiao He
Zidong Wang
Zhijie Zhou
Youqing Wang
Qiang Liu*

Research Article

Deep Sparse Autoencoder for Feature Extraction and Diagnosis of Locomotive Adhesion Status

Changfan Zhang ¹, Xiang Cheng,^{1,2} Jianhua Liu,¹ Jing He ¹ and Guangwei Liu¹

¹College of Electrical and Information Engineering, Hunan University of Technology, Zhuzhou 412000, China

²CHNV Technology Co., Ltd., Dongguan 523000, China

Correspondence should be addressed to Jing He; hejing@263.net

Received 6 January 2018; Revised 30 May 2018; Accepted 27 June 2018; Published 15 July 2018

Academic Editor: Qiang Liu

Copyright © 2018 Changfan Zhang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The model is difficult to establish because the principle of the locomotive adhesion process is complex. This paper presents a data-driven adhesion status fault diagnosis method based on deep learning theory. The adhesion coefficient and creep speed of a locomotive constitute the characteristic vector. The sparse autoencoder unsupervised learning network studies the input vector, and the single-layer network is superimposed to form a deep neural network. Finally, a small amount of labeled data is used to fine-tune training the entire deep neural network, and the locomotive adhesion state fault diagnosis model is established. Experimental results show that the proposed method can achieve a 99.3% locomotive adhesion state diagnosis accuracy and satisfy actual engineering monitoring requirements.

1. Introduction

Precise diagnosis of the wheel-rail adhesion state is an important prerequisite for adhesion control. Currently, the wheel-rail adhesion state of a locomotive is mostly diagnosed based on the detection and analysis of relevant parameters to determine the type of adhesion state and the degree of adhesion [1]. For the diagnosis of adhesion states, a sampling eigenvector should be generated based on the creeping speed of the driving wheel and the wheel-rail cohesion coefficient, a sample feature should be extracted, and the feature should be coded; then, various intelligent algorithms should be used to classify the eigenvector [2]. Many studies on the use of neural networks in the adhesion field have been reported. For example, Castillo et al. [3] used a neural network to estimate the adhesion state in an ABS system. Castillo [4] trained an artificial neural network to calculate the best creep operating point for each road on the basis of traffic information collected by a vehicle sensor. Li Ningzhou [5] studied the adhesion feature of the air brake of a locomotive and used the optimized recursive neural network to optimize the parameters of the adhesive controller and improve the

utilization rate of locomotive adhesion, thereby obtaining a good experimental result.

These methods are more convenient and intelligent than the general mechanism analysis method. However, these methods still belong to the supervised learning area [6]. Thus, they require sufficient data for feature extraction. Meanwhile, extracting the right features is often relatively complex and difficult. To obtain labels, experiments and rich professional knowledge are required. With the artificial participation factors, the uncertainty of feature extraction and optimization is greatly increased, thereby making the diagnosis of the right adhesion state difficult. Furthermore, a traditional neural network essentially uses hidden-layer neurons for nonlinear transformation [7]. It can learn potential features from a given sample and fit out an approximation function [8]. Taking the classical BP neural network as an example, obtaining high-precision features becomes more difficult while the layers are few. If the number of layers is excessive, then the gradient may disappear, and the local optimal solution is another defect that is difficult to overcome [9].

Sample feature extraction is a key step in determining the accuracy of fault diagnosis [10]. The change in adhesion

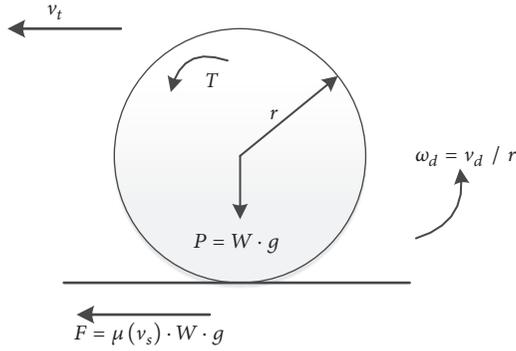


FIGURE 1: Schematic of adhesion.

state is a complex process that is affected by multiple factors, producing a complex nonlinear relation between factors and outcomes. Fault prediction and analysis are particularly challenging. The introduction of deep learning [11] has made a breakthrough in research on high-precision feature extraction. As an unsupervised learning algorithm, deep neural network not only has an excellent feature extraction ability but can also overcome the common problem of obtaining sample labels [12]. Thus, deep neural network has become a popular research area in the field of fault diagnosis [13–15]. This paper proposes a sparse autoencoder deep neural network with dropout to diagnose the wheel-rail adhesion state of a locomotive. This deep neural network can significantly reduce the adverse effect of overfitting, making the learned features more conducive to classification and identification.

The rest of this paper is organized as follows: Section 2 describes the adhesion principle and characteristics. Section 3 describes the principle and process of the deep neural network algorithm. Section 4 discusses the comparative experimental research and result analysis. Section 5 presents the conclusions.

2. Description of Adhesion Status

Adhesion is the ultimate manifestation of locomotive driving force in the wheel-rail relationship and the fundamental motive force for locomotives [16]. The wheel pair rolls forward when subjected to a tangential traction, and the rolling pressure causes deformation between the wheel and the rail. Simultaneously, the gravity of the car body imposed on the rail keeps the contact surface between the wheel and the rail relatively stable. This phenomenon is called adhesion. As shown in Figure 1, the contact point between the wheel and the rail is elastically deformed under the action of the wheel load (P). The wheel rolls forward under the action of the driving torque (T), the original contact surface deformation develops into a new elliptical deformation, and the tractive effort at the wheel rim (F) is generated.

Adhesion coefficient μ is typically defined as the ratio of traction to axle weight (see (1)), where W is the axle weight (kg) and g is the gravitational acceleration (m/s^2).

$$\mu = \frac{F}{W \cdot g} \quad (1)$$

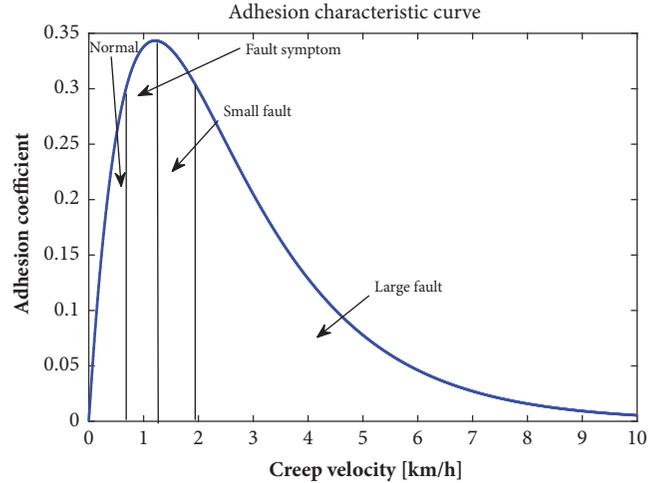


FIGURE 2: Refined adhesion characteristic curve.

In the process of normal movement, the train body speed (v_t) is always less than the wheel speed (v_d) due to the wheel-rail microsliding generated by the deformation. This phenomenon is called creep, and the speed difference between them is defined as creep speed v_s .

$$v_s = v_d - v_t \quad (2)$$

Creep is a slight wheel-spin phenomenon produced by the locomotive drive system. The adhesion coefficient of the rail contact surface rises constantly with the creep speed within a certain range [17], and the locomotive has a great available traction. Once the range is exceeded, the wheel-rail adhesion coefficient drops sharply with the increase in the creep speed.

Figure 2 shows the adhesion characteristic curve of the locomotive. The adhesion peak point is taken as the boundary in which the left side is called the creep region and the right side is called the slid region [4]. However, when the adhesion state is divided into two categories, abnormal adhesion can be identified, but the predicted foundation of potential creep failure cannot be provided. To this end, this paper further refines the adhesion features: normal (N0), fault symptom (N1), small fault (F1), and large fault (F2).

The adhesion state is divided into four categories. When a minor fault is encountered, fault tolerant control methods [18] can be adopted to prevent serious system performance deterioration [1].

3. Deep Neural Network

Unsupervised learning can be used to automatically learn potential features from the samples without labels [19, 20]. This method has a significant advantage when addressing complex problems, such as adhesion state recognition. The sparse autoencoder is an unsupervised algorithm, and this deep neural network can effectively extract the characteristics that reflect the adhesion state [21, 22].

3.1. Sparse Autoencoder. From the structural point of view, the autoencoder is an axisymmetric single hidden-layer

neural network [23]. The autoencoder encodes the input sensor data by using the hidden layer, approximates the minimum error, and obtains the best-feature hidden-layer expression [24]. The concept of the autoencoder comes from the unsupervised computational simulation of human perceptual learning [25], which itself has some functional flaws. For example, the autoencoder does not learn any practical feature through copying and inputting memory into implicit layers, although it can reconstruct input data with high precision. The sparse autoencoder inherits the idea of the autoencoder and introduces the sparse penalty term, adding constraints to feature learning for a concise expression of the input data [26, 27].

For the adhesion state identification of locomotive, k sets of monitoring data $\{x_1, x_2, x_3, \dots, x_n\}$ exist, which are reconstructed into a $N \times M$ data set $\{x(1), x(2), x(3), \dots, x(N)\}$, $x(i) \in R^M$. These data are used as input matrix X . The input data encoded by the automatic encoder are used to construct a mapping relationship. In this paper, the activation function of the autoencoder is sigmoid, which is designed to obtain a better representation of input data: $h(X, W, b) = \sigma(WX + b)$. A sparse penalty term is added to the sparse autoencoder cost function to limit the average activation value of the hidden-layer neuron. Normally, when the output value of a neuron is 1, it is active, and the neuron is inactive when its output value is 0. The purpose of enforcing sparsity is to limit the undesired activation. $a_j(x)$ is set as the j th activation value. In the process of feature learning, the activation value of the hidden-layer neuron is usually expressed as $a = \text{sigmoid}(WX + b)$, where W is the weight matrix and b is the deviation matrix. The mean activation value of the j th neuron in the hidden layer is defined as

$$\rho_j = \frac{1}{n} \sum_{i=1}^n [a_j(x(i))]. \quad (3)$$

The hidden layer is kept at a lower value to ensure that the average activation value of the sparse parameter is defined as ρ , and the penalty term is used to prevent ρ_j from deviating from parameter ρ . The Kullback–Leibler (KL) divergence [28] is used in this study as the basis of punishment. The mathematical expression of KL divergence is as follows:

$$KL(\rho \parallel \rho_j) = \rho \ln \frac{\rho}{\rho_j} + (1 - \rho) \ln \frac{1 - \rho}{1 - \rho_j}. \quad (4)$$

When ρ_j does not deviate from parameter ρ , the KL divergence value is 0; otherwise, the KL divergence value will gradually increase with the deviation. The cost function of the neural network is set as $C(W, b)$. Then, the cost function of adding the sparse penalty term is

$$C_{\text{sparse}} = C(W, b) + \beta \sum_{j=1}^{S_2} KL(\rho \parallel \rho_j) \quad (5)$$

where S_2 is the number of neurons in the implicit layer and β is the weight of the sparse penalty term. The training essence of a neural network is to find the appropriate weight and threshold parameter (W, b). After the sparse penalty term is defined, the sparse expression can be obtained by minimizing the sparse cost function.

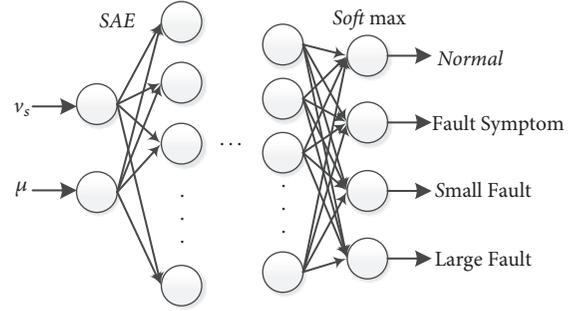


FIGURE 3: Schematic of the network structure.

3.2. Softmax Regression. The sparse autoencoder can form the deep network structure through the multilayer stack, which can be used for feature learning and clustering of the adhesion data collected by the sensor. However, this autoencoder has no ability to classify. Therefore, this paper presents a deep neural network architecture that combines the stacked sparse self-encoder and softmax regression. The schematic of the network structure is shown in Figure 3.

Softmax regression is an extension of the logistic regression model on multiple classifications [29]. The category tag of the logistic regression can only take two values, whereas the softmax tag can take on multiple values [30]. Let us suppose m training samples of adhesion state $\{(x^{(1)}, y^{(1)}), (x^{(2)}, y^{(2)}), \dots, (x^{(m)}, y^{(m)})\}$, $y^{(i)} \in \{1, 2, \dots, n\}$. The hypothesis function is used to estimate the probability value $p(y = j | x)$ for each category j . The softmax output is defined as follows:

$$h_{\theta}(x^{(i)}) = \begin{bmatrix} p(y^{(i)} = 1 | x^{(i)}; \theta) \\ p(y^{(i)} = 2 | x^{(i)}; \theta) \\ \vdots \\ p(y^{(i)} = k | x^{(i)}; \theta) \end{bmatrix} \quad (6)$$

$$= \frac{1}{\sum_{j=1}^k e^{\theta_j^T x^{(i)}}} \begin{bmatrix} e^{\theta_1^T x^{(i)}} \\ e^{\theta_2^T x^{(i)}} \\ \vdots \\ e^{\theta_k^T x^{(i)}} \end{bmatrix}$$

where $\theta_1, \theta_2, \dots, \theta_k \in \mathfrak{R}^{n+1}$ is the model parameter. $1 / \sum_{j=1}^k e^{\theta_j^T x^{(i)}}$ normalizes the probability distribution such that the sum of all probabilities is 1.

3.3. Overfitting and L2 Regularization. L2 regularization is a way of effectively reducing the neural network overfitting [31]. In this study, this method is used to avoid overlearning on features caused by synergies. The basic principles of L2 regularization are as follows:

$$C'_{\text{sparse}} = C_{\text{sparse}} + \alpha \|w\|^2 \quad (7)$$

TABLE 1: Description of adhesion state of locomotive wheel rail.

| Different types of States | State representation |
|---------------------------|-------------------------------------|
| Normal | Running with a low adhesion |
| Fault symptom | Near the peak adhesion coefficient |
| Small fault | Across the peak adhesion point |
| Large fault | Serious slip running unsafe already |

where C_{sparse} is the cost function of the neural network, α is coefficient, and $\|w\|^2$ is the penalty term. The greater the coefficient, the deeper the weight attenuation.

3.4. Framework of the Algorithm. The feature learning ability of the single sparse autoencoder is limited. To construct a model with improved feature extraction capacity, we stacked the sparse autoencoders into a deep structure (SAE). In this process, the output of the upper layer of the encoder is taken as the input of the next layer to achieve a multilearning sample feature. The flowchart of the deep neural network algorithm is shown in Figure 4 and subsequently described.

(1) The Creep Speed (v_s) and Adhesion Coefficient μ Monitored by the Sensor Are Used to Train the Sparse Autoencoder

- (1) The parameters, such as the network learning rate and the dropout parameter, are set; weights W and thresholds b are initialized.
- (2) The number of iterations is set, and the mean activation value ρ_j and the sparse cost function are calculated according to (3)–(5); the network parameters are updated based on the backpropagation (BP) algorithm.

(2) The Deep Neural Network Is Fine-Tuned with a Small Number of Labeled Samples

- (1) With the abovementioned step, the parameters of threshold b and weight W learned from the network are saved.
- (2) The L2 regularization and learning rates are set, and the mean square error is calculated.
- (3) The BP algorithm is used to update the weights of the network and fine-tune the entire network.

(3) The Performance of the Identification Model Is Tested

- (1) The sample size is usually 30% of the total number of samples.
- (2) According to the L2 coefficient, the weight of neural network is attenuated while performing the front-propagation algorithm.
- (3) The output of DNN is compared with the sample labels, and comparative statistics are made.

4. Experimental Research and Analysis

For performance comparison with this dropout-based deep neural network, a BP neural network and an optimized BP

TABLE 2: State type.

| Different types of states | State encoding |
|---------------------------|----------------|
| Normal (N0) | 000 |
| Fault symptom (N1) | 001 |
| Small fault (F1) | 010 |
| Large fault (F2) | 100 |

neural network based on a genetic algorithm (GA-BP) are used as state recognition models. Taking the creep speed (v_s) and the adhesion coefficient (μ) of the locomotive as characteristic signals, the adhesion features include normal operation zone, wheel-spin warning zone, slight wheel-spin zone, and serious wheel-spin. Detailed information is shown in Table 1. The test sample added a Gaussian noise with a mean of 0 and a variance of 0.02 for an improved fit construction.

The changes in the state of adhesion directly affect the running safety of a locomotive. These changes are reflected in the sensor monitoring data. In this section, the adhesion state of the locomotive is identified and diagnosed according to the inherent characteristics of the sample data of deep neural network based on the sparse autoencoder. In general, the original data samples of sensors are divided into training and test sets according to a 7:3 ratio. A total of 700 training samples and 300 test samples are used in this experiment.

4.1. Simulation of the BP Neural Network. We need to set parameters before the experiment. The number of hidden-layer nodes is set according to the empirical formula $l = 2n + 1$, where n is the input dimension and l is the number of hidden-layer nodes. In this study, the creep speed (v_s) and the adhesion coefficient (μ) are used as inputs, so that l should be 5, and the adhesion state is expressed in binary form, as shown in Table 2. This BP neural network should be a $2 \times 5 \times 3$ structure (Figure 5). W is the weight and β is the bias of the BP neural network. The mean square error curve is shown in Figure 6.

To demonstrate the advantages of the proposed algorithm, a genetic algorithm- (GA-) optimized BP neural network is used as a contrast experiment. The crossover operator uses a single point crossover, the crossover probability is 0.7, and the mutation probability is 0.01. The evolutionary process of the GA is shown in Figure 7, and the error descent curve of the BP neural network that is optimized by the GA is shown in Figure 8.

4.2. Simulation of Sparse Autoencoder Deep Neural Network. The visualization of the target classification is shown in Figure 9 to provide a clear analysis of the classification ability of the proposed algorithm. The plane between the yellow and blue modules is the desired classification plane. In this experiment, 1,000 sets of monitoring data are selected as experimental samples, and the 7:3 ratio is used to divide the samples into training and test samples. The actual results are shown in Figure 10, which shows that the actual classification plane is basically consistent with the expected one.

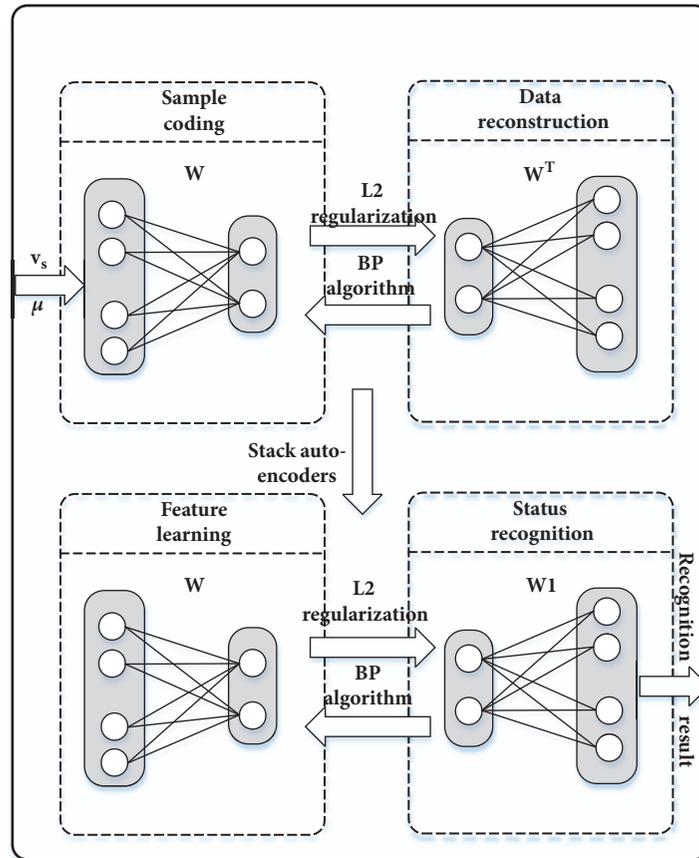


FIGURE 4: Flowchart of deep neural network algorithm.

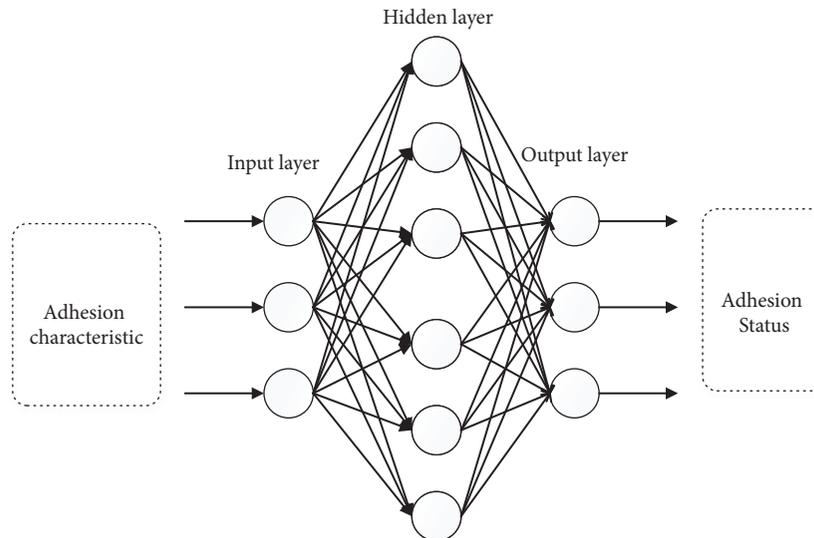


FIGURE 5: Schematic of the structure of the BP neural network.

Figure 9 shows the visualization of the classification target of adhesion status. In general, it is necessary to divide the adhesion status into four different statuses. Three classification planes are needed to achieve this (yellow and blue junction in the figure). The requirement of training

neural network to accurately classify the adhesion state of locomotives is to make the test dataset also present a clear classification plane. The actual state division result is shown in Figure 10. The classifying plane of the adhesion state is clear.

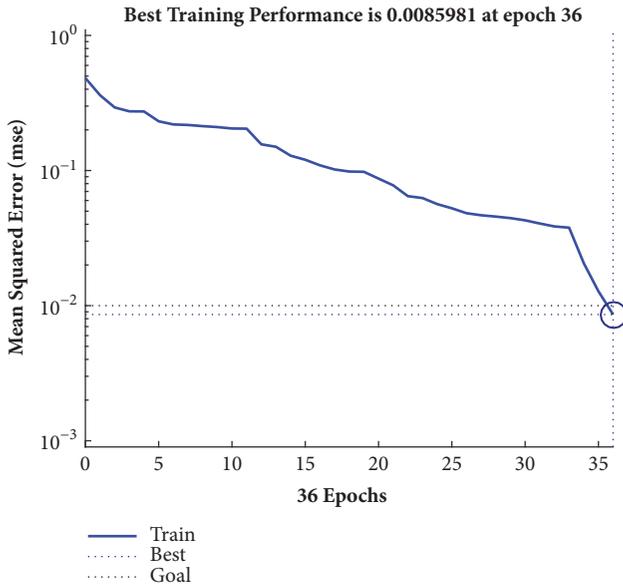


FIGURE 6: Simulation error curves of BP neural network.

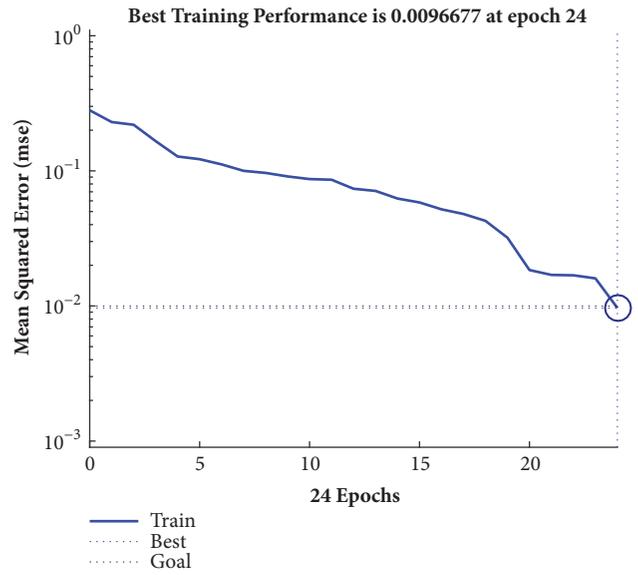


FIGURE 8: Simulation error curves of the GA-BP neural network.

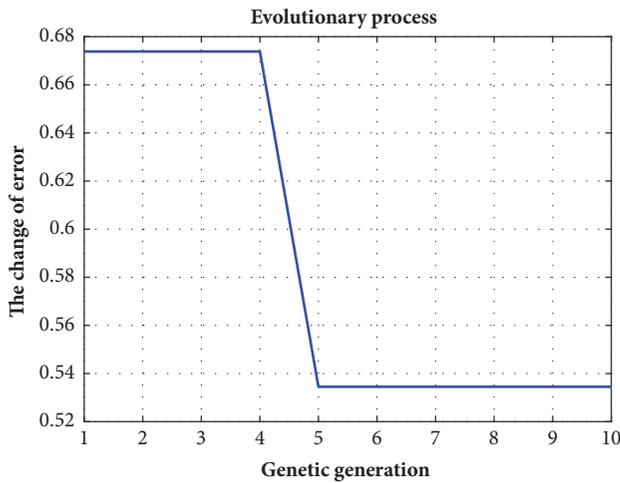


FIGURE 7: Evolutionary process.

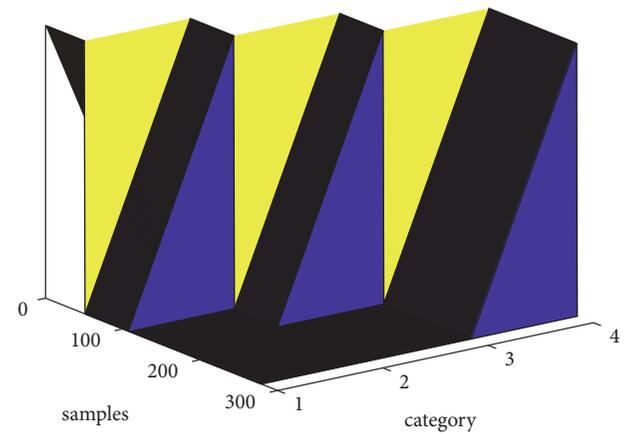


FIGURE 9: Classification target visualization.

From the error histogram in Figure 11, the error distribution of the deep neural network in this chapter is basically in line with the normal distribution, which meets the needs of practical application. The accuracy of adhesion state recognition is shown in Figures 12 and 13. The horizontal axis represents the desired target category, the vertical axis represents the experimentally predicted adhesion state category, and the gray block shows the exact percentage of prediction and expectation.

Figure 12 shows that the accuracy of the deep neural network adhesion state recognition is 96.1%. The overfitting of the neural network generally appears as the trained neural network does not accurately identify the test samples. The adhesion state is a continuously changing process; in order to ensure the safety of driving, the identification of the adhesion state must be as accurate as possible. Since the recognition accuracy rate does not reach the ideal state and there are

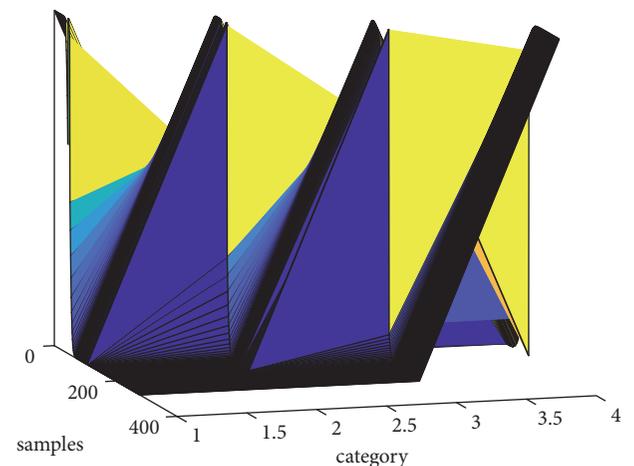


FIGURE 10: Contrast experiment with sparse coefficient 0.08.

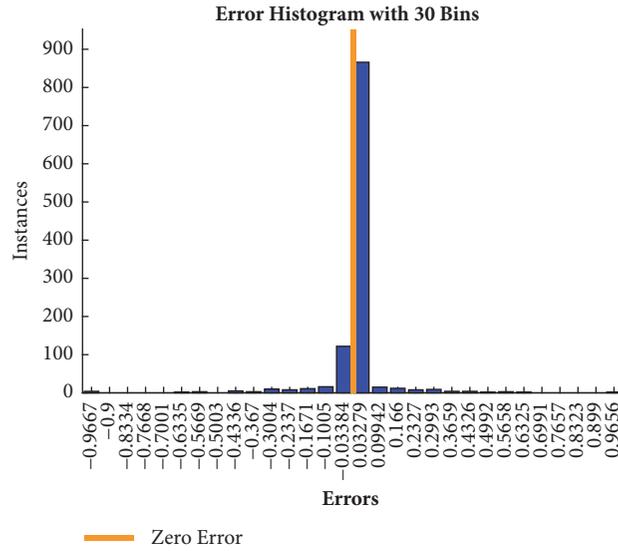


FIGURE 11: Error histogram.

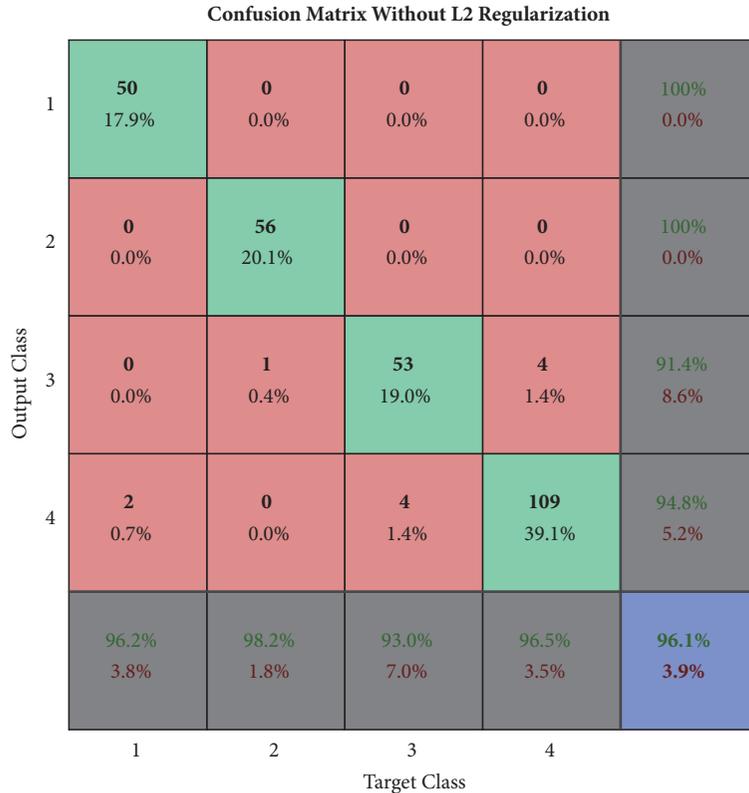


FIGURE 12: Confusion matrix without L2 regularization.

significantly more training samples than test samples, there is ample reason to speculate that the deep neural network used in this section has been overfitted. To improve this phenomenon, the L2 regularization method was used to attenuate the weights of the deep neural network. Figure 13 shows the results of the deep neural network adhesion state recognition after L2 regularization. The accuracy rate reaches

99.6%. The accuracy rate of the neural network for the adhesion state test set is improved, and the proposed L2 regularization can improve the overfitting phenomenon that may occur in the adhesion state recognition based on deep neural network.

The experimental results show that the SAE-based locomotive adhesion diagnosis can meet the requirements of

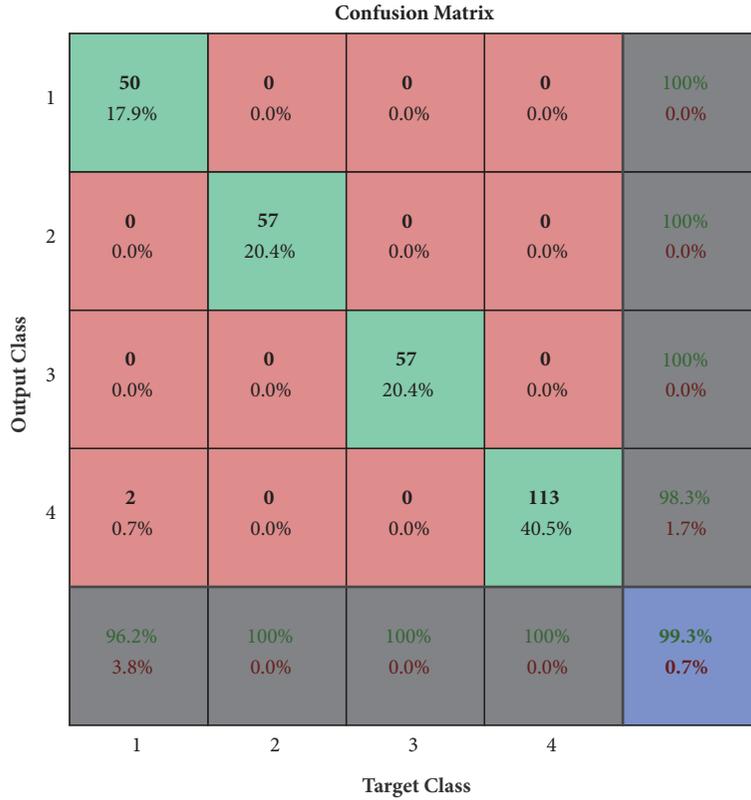


FIGURE 13: Confusion matrix.

TABLE 3: Test accuracy of the four methods.

| Method | Type of adhesion state | | | |
|--------|------------------------|-------|-------|-------|
| | N0 | N1 | F1 | F2 |
| BP | 83.33 | 57.14 | 77.78 | 95.45 |
| GA-BP | 84.23 | 55.49 | 88.62 | 97.66 |
| DNN | 96.2 | 100 | 100 | 99.3 |

high-accuracy recognition under a reasonable error distribution. Table 3 presents the comparison of the three algorithms in this paper with the traditional BP neural network and GA-optimized neural network performance.

5. Conclusions

In this paper, an adhesion state fault diagnose method based on SAE is proposed. The effectiveness of the proposed method is validated by computer simulation. The conclusions are elaborated as follows:

- (1) The adhesion state is divided into four categories, which provide a strong basis for wheel skid warning.
- (2) The sparse automatic encoder can extract data features effectively, make classification easier, and extract more robust data features.
- (3) Compared with the traditional BP neural network, the deep neural network of the sparse autoencoders can ensure effective fault diagnosis of the locomotive adhesion condition.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare no conflicts of interest.

Authors' Contributions

Changfan Zhang and Xiang Cheng conceived and designed the research; Changfan Zhang and Xiang Cheng performed the research; Changfan Zhang, Xiang Cheng, Jing He, Jianhua Liu, and Guangwei Liu wrote the paper.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (nos. 61773159, 61473117), Hunan

Provincial Natural Science Foundation of China (nos. 2018JJ2093, 2017JJ4031), Key Laboratory for Electric Drive Control and Intelligent Equipment of Hunan Province (no. 2016TP1018), and Science and Technology Innovative Research Team in Higher Educational Institutions of Hunan Province.

References

- [1] C. Zhang, J. Sun, J. He, and L. Liu, "Online Estimation of the Adhesion Coefficient and Its Derivative Based on the Cascading SMC Observer," *Journal of Sensors*, vol. 2017, Article ID 8419295, 11 pages, 2017.
- [2] C. Zhang, X. Cheng, J. He, and G. Liu, "Automatic recognition of adhesion states using an extreme learning machine," *International Journal of Robotics and Automation*, vol. 32, no. 2, pp. 194–200, 2017.
- [3] J. J. Castillo, J. A. Cabrera, A. J. Guerra, and A. Simón, "A Novel Electrohydraulic Brake System with Tire-Road Friction Estimation and Continuous Brake Pressure Control," *IEEE Transactions on Industrial Electronics*, vol. 63, no. 3, pp. 1863–1875, 2016.
- [4] J. J. C. Aguilar, J. A. C. Carrillo, A. J. G. Fernández, and E. C. Acosta, "Robust road condition detection system using in-vehicle standard sensors," *Sensors*, vol. 15, no. 12, pp. 32056–32078, 2015.
- [5] N. Li, X. Feng, and X. Wei, "Optimized adhesion control of locomotive airbrake based on GSA-RNN," in *Proceedings of the 7th International Conference on Intelligent Human-Machine Systems and Cybernetics, IHMSC 2015*, pp. 157–161, China, August 2015.
- [6] A. Stanescu and D. Caragea, "An empirical study of ensemble-based semi-supervised learning approaches for imbalanced splice site datasets," *BMC Systems Biology*, vol. 9, no. 5, article no. S1, 2015.
- [7] M. Madhjarasan and S. N. Deepa, "Comparative analysis on hidden neurons estimation in multi layer perceptron neural networks for wind speed forecasting," *Artificial Intelligence Review*, vol. 48, no. 4, pp. 449–471, 2017.
- [8] N. J. Guliyev and V. E. Ismailov, "A single hidden layer feed-forward network with only one neuron in the hidden layer can approximate any univariate function," *Neural Computation*, vol. 28, no. 7, pp. 1289–1304, 2016.
- [9] M. Cai and J. Liu, "Maxout neurons for deep convolutional and LSTM neural networks in speech recognition," *Speech Communication*, vol. 77, pp. 53–64, 2016.
- [10] Y. Liu, Y. Li, X. Ma, and R. Song, "Facial expression recognition with fusion features extracted from salient facial areas," *Sensors*, vol. 17, no. 4, 2017.
- [11] G. E. Hinton, S. Osindero, and Y. Teh, "A fast learning algorithm for deep belief nets," *Neural Computation*, vol. 18, no. 7, pp. 1527–1554, 2006.
- [12] X. Ma, Z. Dai, Z. He, J. Ma, Y. Wang, and Y. Wang, "Learning traffic as images: a deep convolutional neural network for large-scale transportation network speed prediction," *Sensors*, vol. 17, no. 4, 2017.
- [13] Xiao He, Zidong Wang, Gang Li, Zhijie Zhou, and Youqing Wang, "Fault Diagnosis and Application to Modern Systems," *Journal of Control Science and Engineering*, vol. 2017, pp. 1–3, 2017.
- [14] C. Li, R.-V. Sánchez, G. Zurita, M. Cerrada, and D. Cabrera, "Fault diagnosis for rotating machinery using vibration measurement deep statistical feature learning," *Sensors*, vol. 16, no. 6, article no. 895, 2016.
- [15] W. Zhang, G. Peng, C. Li, Y. Chen, and Z. Zhang, "A new deep learning model for fault diagnosis with good anti-noise and domain adaptation ability on raw vibration signals," *Sensors*, vol. 17, no. 2, 2017.
- [16] T. Ishrat, G. Ledwich, M. Vilathgamuwa, and P. Borghesani, "Wheel slip control based on traction force estimation of electric locomotives," in *Proceedings of the Australasian Universities Power Engineering Conference, (AUPEC '16)*, Australia, 2016.
- [17] M. Spiryagin, P. Wolfs, C. Cole et al., "Theoretical investigation of the effect of rail cleaning by wheels on locomotive tractive effort," in *Core*, p. V001T10A003, 2016.
- [18] Xiao He, Yamei Ju, Yang Liu, and Bangcheng Zhang, "Cloud-Based Fault Tolerant Control for a DC Motor System," *Journal of Control Science and Engineering*, vol. 2017, pp. 1–10, 2017.
- [19] H. Elghazel and A. Aussem, "Unsupervised feature selection with ensemble learning," *Machine Learning*, vol. 98, no. 1–2, pp. 157–180, 2013.
- [20] F. Anselmi, J. Z. Leibo, L. Rosasco, J. Mutch, A. Tacchetti, and T. Poggio, "Unsupervised learning of invariant representations," *Theoretical Computer Science*, vol. 633, pp. 112–121, 2016.
- [21] W. Zhou, Z. Shao, C. Diao, and Q. Cheng, "High-resolution remote-sensing imagery retrieval using sparse features by auto-encoder," *Remote Sensing Letters*, vol. 6, no. 10, pp. 775–783, 2015.
- [22] O. Tsinalis, P. M. Matthews, and Y. Guo, "Automatic Sleep Stage Scoring Using Time-Frequency Analysis and Stacked Sparse Autoencoders," *Annals of Biomedical Engineering*, vol. 44, no. 5, pp. 1587–1597, 2016.
- [23] M. Kang, K. Ji, X. Leng, X. Xing, and H. Zou, "Synthetic aperture radar target recognition with feature fusion based on a stacked autoencoder," *Sensors*, vol. 17, no. 1, 2017.
- [24] J. Leng and P. Jiang, "A deep learning approach for relationship extraction from interaction context in social manufacturing paradigm," *Knowledge-Based Systems*, vol. 100, no. C, pp. 188–199, 2016.
- [25] B. A. Olshausen and D. J. Field, "Emergence of simple-cell receptive field properties by learning a sparse code for natural images," *Nature*, vol. 381, no. 6583, pp. 607–609, 1996.
- [26] D. Yang, J. Lai, and L. Mei, "Deep representations based on sparse auto-encoder networks for face spoofing detection," *Springer International Publishing*, 2016.
- [27] J. Xu, L. Xiang, Q. Liu et al., "Stacked sparse autoencoder (SSAE) for nuclei detection on breast cancer histopathology images," *IEEE Transactions on Medical Imaging*, vol. 35, no. 119, 2016.
- [28] A. Youssef, C. Delpha, and D. Diallo, "An optimal fault detection threshold for early detection using Kullback-Leibler Divergence for unknown distribution data," *Signal Processing*, vol. 120, pp. 266–279, 2016.
- [29] J. Yang, Y. Bai, G. Li, M. Liu, and X. Liu, "A novel method of diagnosing premature ventricular contraction based on sparse auto-encoder and softmax regression," *Bio-Medical Materials and Engineering*, vol. 26, pp. S1549–S1558, 2015.
- [30] M. M. A. Rahhal, Y. Bazi, H. Alhichri, N. Alajlan, F. Melgani, and R. R. Yager, "Deep learning approach for active classification of electrocardiogram signals," *Information Sciences*, vol. 345, pp. 340–354, 2016.
- [31] E. Phaisangittisagul, "An Analysis of the Regularization Between L2 and Dropout in Single Hidden Layer Neural Network," in *Proceedings of the 7th International Conference on Intelligent Systems, Modelling and Simulation, ISMS 2016*, pp. 174–179, Thailand, January 2016.

Research Article

Data Preprocessing Method and Fault Diagnosis Based on Evaluation Function of Information Contribution Degree

Siyu Ji  and Chenglin Wen 

School of Automation, Hangzhou Dianzi University, Hangzhou 310018, China

Correspondence should be addressed to Chenglin Wen; wenc1@hdu.edu.cn

Received 18 February 2018; Revised 14 April 2018; Accepted 30 April 2018; Published 2 July 2018

Academic Editor: Youqing Wang

Copyright © 2018 Siyu Ji and Chenglin Wen. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Neural network is a data-driven algorithm; the process established by the network model requires a large amount of training data, resulting in a significant amount of time spent in parameter training of the model. However, the system modal update occurs from time to time. Prediction using the original model parameters will cause the output of the model to deviate greatly from the true value. Traditional methods such as gradient descent and least squares methods are all centralized, making it difficult to adaptively update model parameters according to system changes. Firstly, in order to adaptively update the network parameters, this paper introduces the evaluation function and gives a new method to evaluate the parameters of the function. The new method without changing other parameters of the model updates some parameters in the model in real time to ensure the accuracy of the model. Then, based on the evaluation function, the Mean Impact Value (MIV) algorithm is used to calculate the weight of the feature, and the weighted data is brought into the established fault diagnosis model for fault diagnosis. Finally, the validity of this algorithm is verified by the example of UCI-Combined Cycle Power Plant (UCI-ccpp) simulation of standard data set.

1. Introduction

In this paper, based on the evaluation function to calculate the weight of the data feature, the premise is to evaluate the accuracy of the function, which is the theoretical basis and guarantee for the validity of the follow-up work. At present, BP neural network is one of the most popular regression algorithms. It can effectively approximate complex nonlinear mappings based on inputting samples and has the advantages of simple structure and high operability and has been applied in many fields [1, 2]. On the other hand, there are some problems [3]. First of all, the parameter training method based on the gradient descent learning algorithm converges slowly and falls into the local optimum. Secondly, there are many parameters in the neural network that need to be trained, which can take a great time to deal with. In the actual operation of the system, the model data of the system may not have been completely collected but acquired one by one or block by block, while the neural network is a learning algorithm lacking the ability to update online. It

is unacceptable for many practical situations to retrain the network in response to changes in the modalities of system. However, if the parameters in the neural network are not updated in time, the fitting output of the network will greatly deviate from the real value.

In order to solve the above problems, the industry has proposed Neural Network Incremental Learning algorithm to deal with them [4]. The incremental learning method can adjust the artificial neural network by analyzing the specific conditions and recognition results of each new sample, learn new knowledge based on the existing knowledge, and flexibly adapt to the dynamic changes of the environment [5]. Therefore, there are many researches on incremental learning algorithms [6, 7]. The main idea of incremental learning is mainly reflected in two aspects: (1) in the actual perception of data, the amount of data is often gradually increased. Therefore, in the face of new data, the learning method should only update the changes caused by the new data without changing the original knowledge base and then learn from the new data contained in the knowledge; (2) the

cost of modifying a well-trained system is usually less than the cost of retraining one system. There are many frameworks for incremental learning. The core of each framework is to evaluate the similarity between new data and stored knowledge. The method thus determines the way in which new knowledge is perceived and the knowledge base is increased, which affects the growth of knowledge. Therefore, the judgment mechanism of new knowledge is the core part of incremental learning. In an Orthogonal Least Square (OLS) learning algorithm proposed by Chen, structure and parameter identification are performed simultaneously [8]. In [9], a dynamic fuzzy neural network (DFNN) based on RBF is proposed. The parameters are adjusted by Linear Least Square (LLS), and the structure can be adaptively added and deleted according to the rules. On the basis of these, a generalized DFNN (GDFNN) is proposed by the literature [10]. Based on the Ellipse Basis Function (EBF), an online parameter locating mechanism is proposed. The above algorithms are more difficult to achieve and at the same time cannot guarantee the real-time algorithm.

The Kalman filtering method was widely used in the fields of process control, communication, biomedical science, etc., once proposed in the 1960s. Because of its recursive nature, it does not need to store a large amount of historical information and reduces the amount of computer storage. Combining the system state equation and the observation equation directly, it can directly give the estimation accuracy when estimating system state parameters. Its concise way of thinking had become the theoretical basis for the development of such theories as estimation theory and emerging information fusion [11, 12]. Compared with Kalman, the least squares method uses all the observed data to estimate the value of the state quantity at the initial time. Due to the large number of observations and the statistical characteristics of the method, then this method has a higher accuracy. However, due to its centralized nature, it lacks real-time performance. Kalman filter after the observation data is updated, the state variables are improved with new observation data to obtain the state variables at this observation time. Therefore, Kalman filtering is suitable for real-time processing. In this paper, a method of real-time update of hidden layer output weights based on Kalman filter is proposed, which avoids the retraining of the model. At the same time, the deviation from the real value of the model output data caused by the failure of the model parameters to be updated due to the system modal change is eliminated.

The rest of this article is organized as follows: Section 2 gives a brief description of the problems to be solved. The formal description of BP neural network, the gradient descent method, least squares method, and Kalman filter method are used to update the global or local parameters of the network in Section 3. In Section 4, we give a detailed introduction to the process of weighting MIV algorithm. Section 5 simulates the proposed algorithm based on the UCI standard dataset and results of comparisons and analyses. Section 6 summarizes the related research contents, elaborates on the existing problems, and gives a prospect of the next work.

2. Problem Description

Based on the data-driven fault diagnosis method, a series of data processing is often required before the fault diagnosis, such as data standardization, data dimension reduction, feature selection, feature weighting, etc. Some even need to map the data onto high-dimensional space, like we are using support vector machine (SVM) for data classification. The ultimate goal of all the above data preprocessing operations is to improve the diagnostic performance of the fault diagnosis method. Different data preprocessing operations are often adopted for different fault diagnosis methods. Feature weighting algorithm is relatively special, and no matter which kind of data preprocessing operation, you can then weight their features. The purpose of the weight of the feature is to amplify the difference between the feature variables of the data and to eliminate the feature redundancy to a certain extent.

$X = [x(1), x(2), \dots, x(N)]$ is the data set composed of N samples and $x(k) = [x_1(k), x_2(k), \dots, x_n(k)]^T$ is a sample of the data set X , where $x(k)$ is the sample at time k and $x_i(k)$ is the component at time k , $k = 1, 2, \dots, N$; $i = 1, 2, \dots, n$. Our purpose is to find a transformation matrix P to perform a weighted transformation on the sample data

$$\bar{x}(k) = Px(k) \quad (1)$$

$$\bar{x}(k) = \begin{bmatrix} p_1 & & & \\ & p_2 & & \\ & & \ddots & \\ & & & p_n \end{bmatrix} \begin{bmatrix} x_1(k) \\ x_2(k) \\ \vdots \\ x_n(k) \end{bmatrix} = \begin{bmatrix} p_1 x_1(k) \\ p_2 x_2(k) \\ \vdots \\ p_n x_n(k) \end{bmatrix} \quad (2)$$

Making the weighted transformed sample data $\bar{x}(k)$ can be more effective in fault diagnosis. The next question is how to model and find the transformation matrix P : the traditional principal component analysis (PCA) algorithm can solve the transformation matrix to map the original data containing n features onto the selected m -dimensional feature space and have $m < n$, so PCA lost some information in the process of data transformation. The information entropy feature weighting algorithm weights the sample data based on the uncertainty of the feature set's classification of the data set, but its calculation is more difficult; the data set also has higher requirements. In this paper, we want to obtain the weight of the feature through an evaluation function. The evaluation function is generated based on the BP neural network fitting function. In order to ensure the accuracy of the evaluation function, we generally need to retrain the network when the system modal changes. This is a time-consuming task. In order to avoid the retraining of the model, this paper proposes a Kalman filter-based algorithm for real-time update of hidden layer output weights of BP neural network and then establishes the evaluation function model to obtain the transformation matrix (i.e., feature weights). Following we are going to introduce the establishment of the evaluation function and the weight of the data feature in detail. The algorithm implementation process is shown in Figure 1.

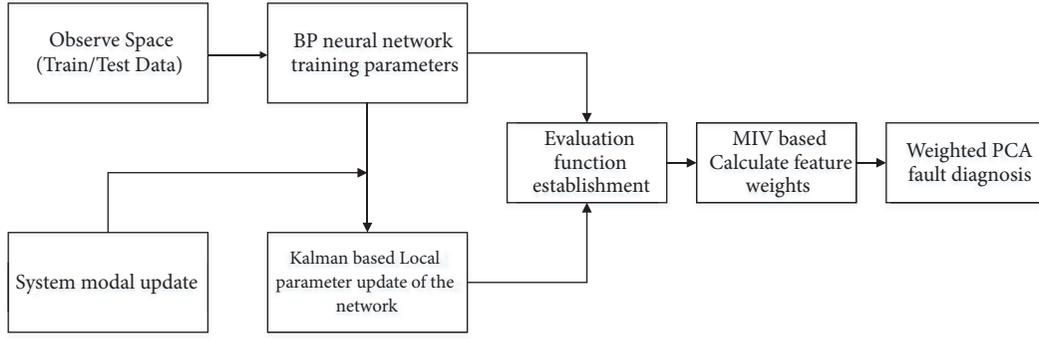


FIGURE 1: Evaluation function of information contribution degree algorithm process.

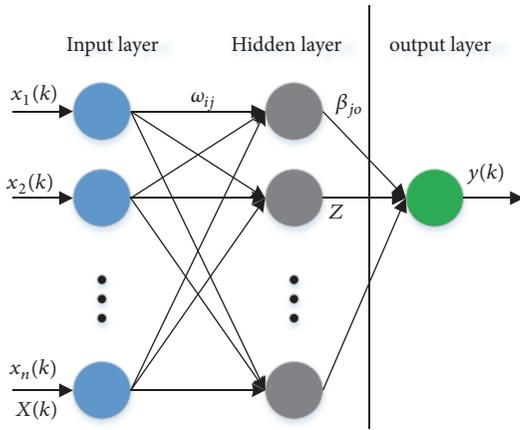


FIGURE 2: BP neural network structure.

3. Evaluation Function Established

The weight of the feature in the above problem depends on the evaluation function. The evaluation function model is generated based on the fitting function of the neural network (i.e., the algebraic representation of the model), and the influence of the input to the output is calculated by the Mean Impact Value (MIV) algorithm, so this article refers to it as the evaluation function. In order to ensure the accuracy of the evaluation function, a precise system model will be necessary. Next, we will conduct a modeling analysis on how to build a model that is accurate and the model parameters updated in real time as the system modal changes.

3.1. Formal Description of BP Neural Network. BP neural network is a kind of feedforward neural network; its main feature is the signal forward, the error of the reverse transmission. In the forward transfer process, the signal input is processed by the hidden layer and finally output. If the output layer error is greater than our set value, it is propagated backwards. After this is repeated, constantly adjust the network weights and bias until the output error meets the set value, or the number of iterations of the algorithm goes up to the ceiling [13]. BP neural network structure is shown in Figure 2.

For the dataset $X = [x(1), x(2), \dots, x(N)]$, we need corresponding output $Y = [y(1), y(2), \dots, y(N)]$ in neural

network training and bring the above labeled dataset into the established BP neural network structure model as (3) to train the model parameters and get the parameters to estimate ω, a, β, b

$$y(k) = \sum_{j=1}^l \beta_{jo} f \left(\sum_{i=1}^n \omega_{ij} x_i(k) + a_j \right) + b, \quad (3)$$

$$k = 1, 2, \dots, N$$

where ω and β are input layer output weight matrix and hidden layer output weight matrix, a and b are the hidden layer bias and output layer bias vector, $\omega \in R^{l \times n}$, $\beta \in R^{l \times m}$ and $a \in R^{l \times 1}$, $b \in R^{m \times 1}$, respectively. BP neural network can be regarded as a nonlinear function, neural network output value, and predicted value of the function of the independent variables and dependent variables. When the input vector is n -dimensional and the output vector is m -dimensional, the BP neural network expresses the function mapping from n -dimensional space onto m -dimensional space.

BP neural network conducts network training before making predictions; the training process includes the following steps.

Step 1. Initialization of neural network parameter: according to the input and output sequence of the system, the number of hidden layer nodes l and the number of output layer nodes m are determined. The weights ω and β of the input layer to the hidden layer and the hidden layer to the output layer are initialized, and the bias vectors are a and b . Set the learning rate η and select the activation function.

Step 2. Calculating output of hidden layer: according to the input of input layer, input layer to the hidden layer of the weight, and bias, calculate output of hidden layer z

$$z_j(k) = f \left(\sum_{i=1}^n \omega_{ij} x_i(k) + a_j \right), \quad (4)$$

$$i = 1, 2, \dots, n; j = 1, 2, \dots, l$$

where f is the hidden layer of the excitation function and the function has many forms, common like sigmoid function, and Gaussian function.

Step 3. Calculating output of output layer z_{jk} : according to output of hidden layer β , output weight of hidden layer vector a , and bias vector b , we calculate the BP neural network prediction output value \hat{y}

$$\hat{y}(k) = \sum_{j=1}^l \beta_{jo} f \left(\sum_{i=1}^n \omega_{ij} x_i(k) + a_j \right) + b, \quad (5)$$

$$o = 1, 2, \dots, m$$

Step 4. Error calculation: based on the predicted output $\hat{y}(k)$ of system and expected output $y(k)$, a network prediction error e is calculated

$$e = \sum_{k=1}^L [\hat{y}(k) - y(k)] \quad (6)$$

Step 5. Updating of weight vector and bias vector: we are going to describe in detail the different weight update algorithms in the following section.

Step 6. Determine whether $\|e(k)\|$ is smaller than the set value or whether the number of iterations of the algorithm has reached the ceiling. If it is “yes” the algorithm ends; if it is “no” the algorithm returns to Step 2.

3.2. Estimation of Evaluation Function Parameters Based on Gradient Descent Method. The method of parameter training in BP neural network generally adopts the method of gradient descent. Gradient descent method is used to find the minimum value of the function, and it is an iterative algorithm, which has been widely used in machine learning. Gradient is a vector, the direction of the derivative is a scalar, the direction of the gradient refers to the direction of the maximum direction of the derivative, and gradient modulus is equal to the maximum value of the directional derivative. Solving the gradient is to find the partial derivative for each independent variable, and then the partial derivative as the direction of the independent variable coordinates can be

$$\nabla J(\omega, a, \beta, b) = \left(\frac{\partial J(\omega, a, \beta, b)}{\partial \omega_{ij}}, \dots, \frac{\partial J(\omega, a, \beta, b)}{\partial a_j}, \dots, \frac{\partial J(\omega, a, \beta, b)}{\partial \beta_{jo}}, \dots, \frac{\partial J(\omega, a, \beta, b)}{\partial b_o} \right) \quad (7)$$

Firstly, select an initial point and calculate the gradient of the point and then update the independent variable in the direction of the gradient. If the k -th iteration value of ω_{ij} is $\omega_{ij}^{(k)}$, then

$$\omega_{ij}^{(k+1)} = \omega_{ij}^{(k)} - \eta \nabla_{\omega} J(\omega_{ij}^{(k)}, a, \beta, b) \quad (8)$$

where η is called the step length or the learning rate, which indicates the size of the argument in each iteration step.

In the same way, other parameters in the network iteratively update the expression

$$a_j^{(k+1)} = a_j^{(k)} - \eta \nabla_a f(\omega, a_j^{(k)}, \beta, b) \quad (9)$$

$$\beta_{jo}^{(k+1)} = \beta_{jo}^{(k)} - \eta \nabla_{\beta} f(\omega, a, \beta_{jo}^{(k)}, b) \quad (10)$$

$$b_o^{(k+1)} = b_o^{(k)} - \eta \nabla_b f(\omega, a, \beta, b_o^{(k)}) \quad (11)$$

Note 1. The gradient descent method slows down when approaching the minimum value. There is no standardized way to set the iteration step size of the algorithm. The algorithm may fall into the local optimum and may not be able to obtain the global minimum. It is difficult to update the parameters in the neural network adaptively with the increasing of input samples.

3.3. Estimation of Evaluation Function Parameters Based on Extreme Learning Machine Method. Extreme Learning Machine (ELM) is an algorithm proposed by Professor Huang Guangbin from Nanyang Technological University to solve neural networks. The biggest characteristic of ELM compared to the traditional neural network, especially single hidden neural network, is faster than the traditional learning algorithm.

When we randomly initialize ω and a in BP neural network and set the bias vector b to zero and consider only β , the BP neural network can be regarded as an ELM [14]. The matrix representation of the ELM combined with the above equation (2) is as follows:

$$Z \cdot \beta = Y \quad (12)$$

where Z is the output of the hidden layer node, β is the output weight, and Y is the expected output.

Once the input weight ω_{ij} and the hidden layer bias a_j are given randomly in the ELM algorithm, the output matrix Z of the hidden layer is uniquely determined. A single hidden layer neural network that had been trained can be transformed into solving a linear system: $Z \cdot \beta = Y$. Then the problem of solving the hidden layer output weight vector β is transformed into the least square problem [15].

Note 2. Single hidden layer neural network in the process of iterative algorithm training all the parameters in the network should be retrained, so its training speed is slower. If using extreme learning machine to solve the network parameters, the training speed of the network will be greatly improved. The BP neural network has been reduced to ELM to reduce the computational complexity. However, ω and a are randomly initialized in the network, which leads to unstable output of the network.

Least squares method is a mathematical optimization technique that finds the best match for the data by minimizing the square of the error. The least squares method can be used to find the unknown data easily and minimize the sum of squares of the errors between the obtained data and the

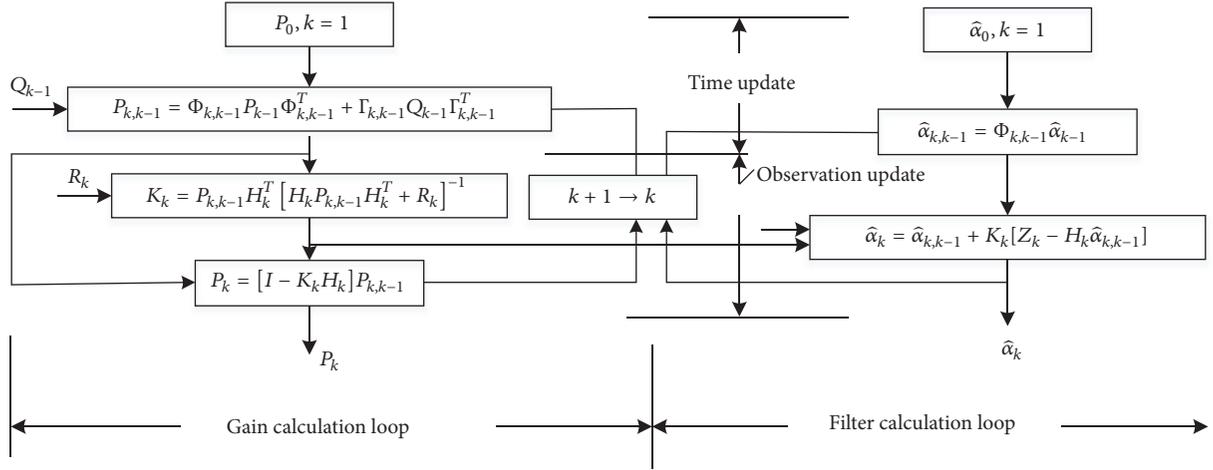


FIGURE 3: The process of Kalman filtering.

actual data. For the linear model established in (11), for the least squares, the final matrix representation is as follows:

$$\min \|Z\beta - Y\|_2 \quad (13)$$

The final optimal solution is

$$\beta = (Z^T Z)^{-1} Z^T Y \quad (14)$$

Note 3. The least squares method is a centralized approach that is based on the whole and cannot adjust and update the network parameters adaptively as the sample data increases.

According to the research, Kalman filtering possesses the potential to adjust parameters in the network in real time as the number of samples increases or modalities update.

3.4. Estimation of Evaluation Function Parameters Based on Kalman. If we want to use Kalman filtering to update the network parameters in real time, then we must establish the state equation and the measurement equation that fit the Kalman filter [16, 17]. Considering that the parameters β and b to be estimated are all subject to a certain degree of random disturbance, we model the state equation of Kalman filter as follows:

$$\alpha(k+1) = \alpha(k) + w(k) \quad (15)$$

In the above formula, $\alpha = [\beta, b]^T$. To simulate the interference to be estimated, we added the white noise sequence $w(k)$ to the equation.

In the above BP neural network algorithm derivation process, (3) can be seen as Kalman filter measurement equation

$$y(k) = \sum_{j=1}^l \beta_{j0} f \left(\sum_{i=1}^n \omega_{ij} x_i(k) + a_j \right) + b + v(k) \quad (16)$$

Considering that we are building a linear Kalman filter model, we randomly initialize w and e in the above formula, so the above equation is distorted to

$$y(k) = H(k) \alpha + v(k) \quad (17)$$

where $H(k) = [g(\sum_{i=1}^n \omega_{i1} x_i(k) + a_1), g(\sum_{i=1}^n \omega_{i2} x_i(k) + a_2), \dots, g(\sum_{i=1}^n \omega_{ij} x_i(k) + a_j), 1]$; since the parameter α to be estimated is slowly changing, we also added the white noise sequence $w(k)$ to the model.

Therefore, the establishment of Kalman filter model is as follows:

$$\alpha(k+1) = A(k+1, k) \alpha(k) + w(k) \quad (18)$$

$$y(k) = H \alpha(k) + v(k)$$

In the Kalman filter model, both the process noise $w(k)$ and the observed noise $v(k)$ are white noise sequences, which are constant values during the sampling interval. With $E\{w(k)w'(k)\} = Q$ and $E\{v(k)v'(k)\} = R$, and when $w(k)$ and $v(k)$ are independent of each other, $E\{w(k)v'(k)\} = 0$. The process of Kalman filtering is shown in Figure 3. Now Kalman optimal filtering equation is summarized as follows:

Time update is

$$\begin{aligned} \hat{\alpha}(k+1|k) &= A(k+1, k) \hat{\alpha}(k|k) \\ &+ B(k+1, k) u(k) \end{aligned} \quad (19)$$

$$\begin{aligned} P(k+1|k) &= A(k+1, k) P(k|k) A^T(k+1|k) \\ &+ \Gamma(k+1, k) Q(k) \Gamma^T(k+1, k) \end{aligned} \quad (20)$$

Measurement update is

$$\begin{aligned} K(k+1) &= P(k+1|k) H^T(k+1) \\ &\cdot [H(k+1) P(k+1|k) H^T(k+1) \\ &+ R(k+1)]^{-1} \end{aligned} \quad (21)$$

$$\hat{\alpha}(k+1|k+1) = \hat{\alpha}(k+1|k) + K(k+1) \cdot [C(k+1)\hat{\alpha}(k+1|k) + v(k+1)] \quad (22)$$

$$P(k+1|k+1) = P(k+1|k) - P(k+1|k)H^T(k+1) \cdot [H(k+1)P(k+1|k)H^T(k+1) + R(k+1)]^{-1} \cdot H(k+1)P(k+1|k) \quad (23)$$

In the above five formulas, (19) to (23), $\hat{\alpha}(k+1|k)$ denotes the state estimation value at time $k+1$; $P(k+1|k)$ denotes the covariance matrix corresponding to $\alpha(k+1|k)$; $K(k+1)$ denotes the optimal gain matrix at time $k+1$; $\hat{\alpha}(k+1|k)$ denotes the state estimation error at time $k+1$, $\hat{\alpha}(k+1|k) = \alpha(k+1) - \hat{\alpha}(k+1|k)$; $\hat{\alpha}(k+1|k+1)$ is the optimal state estimate at time $k+1$; and $P(k+1|k+1)$ represents the covariance matrix corresponding to $\alpha(k+1|k+1)$.

After the above Kalman filter model is established, the observation matrix H is solved with the parameters in the trained neural network, and then the hidden layer output weight vector and the bias vector α are updated according to the Kalman filter algorithm.

Note 4. If we consider ω and a in $f(\sum_{i=1}^n \omega_{ij}x_i(k) + a_j)$, the measurement equation of Kalman filter we have established becomes a nonlinear equation. The equation of state is a column vector $[\omega, a, \beta, b]^T$ consisting of ω , a , β , and b . Nonlinear filtering method for state estimation will greatly increase the complexity of the algorithm [18, 19].

4. Feature Weight Calculation and Fault Diagnosis

Mean Impact Value (MIV) is used to evaluate the importance of each independent variable for the dependent variable size [20]. MIV is an indicator used to determine the magnitude of the effect of input neurons on the output neurons, whose symbols represent the relative direction, and the absolute value represents the relative importance of the effect. In a word, the greater the feature weight calculated by the evaluation function, the more sensitive the output of the function to the change of the characteristic variable. The feature weighting used for sample attributes is mainly to amplify the changes of key feature variables, making the output more sensitive to changes. The specific calculation process is as follows:

$$X_{\pm\delta}^{(i)} = [x_{\pm\delta}^{(i)}(1), x_{\pm\delta}^{(i)}(2), \dots, x_{\pm\delta}^{(i)}(L)] \quad (24)$$

$$x_{\pm\delta}^{(i)}(k) = [x_1(k), \dots, (1 \pm \delta)x_i(k), \dots, x_n(k)]^T \quad (25)$$

$$Z_j = f\left(\sum_{i=1}^n \omega_{ij}x_{\pm\delta}^{(i)}(k) + a_j\right) \quad (26)$$

$$\hat{Y}_{i,\pm} = \sum_{k=1}^L Z_j \beta_{j0} + b_0 \quad (27)$$

$$\begin{aligned} IV_i &= \hat{Y}_{i,+} - \hat{Y}_{i,-} \\ &= F(x_1(k), \dots, (1+\delta)x_i(k), \dots, x_n(k)) \\ &\quad - F(x_1(k), \dots, (1-\delta)x_i(k), \dots, x_n(k)) \end{aligned} \quad (28)$$

Let $\delta \cdot x_i(k) = \Delta x_i$; then (28) becomes

$$\begin{aligned} IV_i &= \hat{Y}_{i,+} - \hat{Y}_{i,-} \\ &= F(x_1(k), \dots, x_i(k) + \delta \cdot x_i(k), \dots, x_n(k)) \\ &\quad - F(x_1(k), \dots, x_i(k) - \delta \cdot x_i(k), \dots, x_n(k)) \quad (29) \\ &= F(x_1(k), \dots, x_i(k) + \Delta x_i, \dots, x_n(k)) \\ &\quad - F(x_1(k), \dots, x_i(k) - \Delta x_i, \dots, x_n(k)) \end{aligned}$$

In summary, take $0.1 \leq \delta \leq 0.3$, $i = 1, 2, \dots, n$. After the network training is terminated, the training samples $X_{+\delta}^{(i)}$ and $X_{-\delta}^{(i)}$ are simulated as the simulation samples using the built-up network to simulate, respectively. Getting two simulation results $\hat{Y}_{i,+}$ and $\hat{Y}_{i,-}$, IV_i are the average influence of the i -th variable in the sample data. Similarly, the average effect of the other variables in the sample data can be obtained.

$$IV = [IV_1, IV_2, \dots, IV_n]^T \quad (30)$$

The parameters in IV are the value of influence for feature variables in sample data to the output. Finally, the IV value of m group is simulated, and the average value of the corresponding mean IV is to be taken as MIV .

$$MIV = \frac{1}{m} \sum_{i=1}^m IV_i \quad (31)$$

The magnitude of the absolute value of MIV is the relative importance of the influence of the respective variables on the network output, thus achieving the weighting of the sample data characteristics, and then the PCA algorithm is used for fault diagnosis.

5. Simulation of the Algorithm

5.1. Simulation Data. The dataset contains 9568 data points collected from a Combined Cycle Power Plant over 6 years (2006-2011), when the power plant was set to work with full load. Features consist of hourly average ambient variables, temperature (T), ambient pressure (AP), relative humidity (RH), and exhaust vacuum (V) to predict the net hourly electrical energy output (EP) of the plant. A Combined Cycle Power Plant (CCPP) is composed of gas turbines (GT), steam turbines (ST), and heat recovery steam generators. In a CCPP, the electricity is generated by gas and steam turbines, which are combined in one cycle, and is transferred from one turbine to another. While the vacuum is collected from and has effect on the steam turbine, the other three of the ambient variables affect the GT performance [21].

In this experiment, 300 sets of data T were used in the simulation test. In the evaluation function establishment

TABLE 1: Neural network fitting results before and after the parameter update.

| Original Data T | The error rate algorithm fitting (%) | | | |
|-----------------|--------------------------------------|-------|-------|-------|
| | 0.99T | 1.00T | 1.02T | 1.05T |
| BPNN | 1.33 | 0.88 | 2.16 | 4.86 |
| Kalman-BPNN | 0.88 | 0.89 | 0.93 | 1.01 |

TABLE 2: The characteristic weights of the sample data.

| Original Data T | Weight | | | | | | | |
|-----------------|-----------------|-------|------|-------|----------------|-------|------|-------|
| | Original weight | | | | Updated weight | | | |
| 0.99T | -10.76 | -0.59 | 0.10 | -3.21 | -7.46 | -1.98 | 0.97 | -1.06 |
| 1.00T | -10.76 | -0.59 | 0.10 | -3.21 | -7.54 | -2.00 | 0.98 | -1.07 |
| 1.02T | -10.76 | -0.59 | 0.10 | -3.21 | -7.69 | -2.03 | 1.00 | -1.09 |
| 1.05T | -10.76 | -0.59 | 0.10 | -3.21 | -7.91 | -2.09 | 1.03 | -1.12 |

and fault diagnosis process, 200 sets were selected as the training data and 100 sets as the test data. To simulate the variation of the system modal, we add an interference $r = 2 + 5 * randn(k, 1)$ to the third feature input of the original dataset, where k is the number of samples. We multiply the original sample data output T by 0.99, 1.02, and 1.05 to simulate changes in the system modality and then test the fitting output of the BP neural network before and after the parameter update. The error rate is used to characterize the fitting accuracy of the model.

Error rate is

$$s_e = \frac{\sqrt{\sum_{i=1}^L (y_i - \hat{y}_i)^2 / n}}{\bar{y}} \times 100\% \quad (32)$$

5.2. Discussion. For the data before and after the system modal change, we use the original trained neural network and the neural networks of hidden layer output weights were updated by Kalman filter algorithm to fit the data, and then we calculated the error rate to characterize the fitting accuracy, respectively. Combined with Figure 4 and Table 1, we can clearly see that the algorithm proposed in this paper real-time update of the model parameters can be very good to ensure the accuracy of the model.

After the traditional neural network parameter training is completed, if the modality of the original system changes, we often retrain the entire network. However, we all know that neural network parameter training process is time-consuming. For those systems whose modality is constantly changing, it is obviously unreasonable to adopt a method of retraining the entire network parameters. The Kalman filtering-based real-time update algorithm for the hidden layer output weights of the BP neural network solves the above problems. On the basis of not changing the other training parameters of the original network, the Kalman filtering model is established, and then the hidden layer output parameters of the BP neural network are updated according to the new data. It can be seen from Table 1 and Figure 4 in this paper that the greater the change in the modality of the original system, the greater the network fitting error. Among the several modal changes we simulated,

the fitting errors were 1.33%, 0.88%, 2.16%, and 4.86%, respectively. It can be seen that the change of the original system modality and the original network can no longer trace the output well. However, in our simulation of several modal changes, using the proposed BP neural network local parameter update algorithm to update the hidden layer output weights in the network, the fitting error is 0.88%, 0.89%, 0.93%, and 1.01%, respectively. It can be seen that using the proposed BP neural network local parameter real-time updating algorithm greatly ensures the accuracy of the model, which is the premise to ensure the validity of the evaluation function.

Table 2 is the vector of weights obtained before and after updating the network parameters. We use the weight vectors obtained before and after updating the network parameters to weight the original sample data and then using the PCA fault diagnosis algorithm to do the fault diagnosis. Finally, the fault diagnosis results of the two kinds of weighted data are compared with the results of failure diagnosis of the sample data without weighting. As shown in Table 3 and Figure 5, the three are significantly different in the SPE statistics. After the modal changes, the weighted weights using the original weights are lower than the unweighted diagnostic accuracy. The weighted weights using the updated weights are significantly higher than the unweighted diagnostic accuracy, with the maximum increase of 18%.

It can be seen from the above simulation data and simulation plot that using Kalman filter algorithm to update the hidden layer output weights of neural networks in real time is feasible. After the observation data is updated, the Kalman filter improves the state quantity with new observation data to get the state quantity at this moment, and the real-time performance of the algorithm is greatly improved. When the system modal changes, the output of the model can be well fitted to the true value without the need of retraining all the parameters of the neural network. At the same time, an accurate evaluation function to obtain the characteristic weight effectively enhances the diagnostic performance of the fault diagnosis algorithm. However, since the method eventually simplifies the model to a linear model and updates some parameters in the network, it is worthwhile to study

TABLE 3: Accuracy of fault diagnosis for different weights.

| Original Data T | Accuracy of fault diagnosis (%) | | | | | |
|-----------------|---------------------------------|-----|-----------------|-----|----------------|-----|
| | Unweighted | | Original weight | | Updated weight | |
| | T^2_{99} | SPE | T^2_{99} | SPE | T^2_{99} | SPE |
| 0.99T | 98 | 50 | 4 | 42 | 98 | 66 |
| 1.00T | 100 | 78 | 100 | 100 | 100 | 84 |
| 1.02T | 98 | 50 | 4 | 42 | 98 | 68 |
| 1.05T | 98 | 50 | 4 | 42 | 98 | 68 |

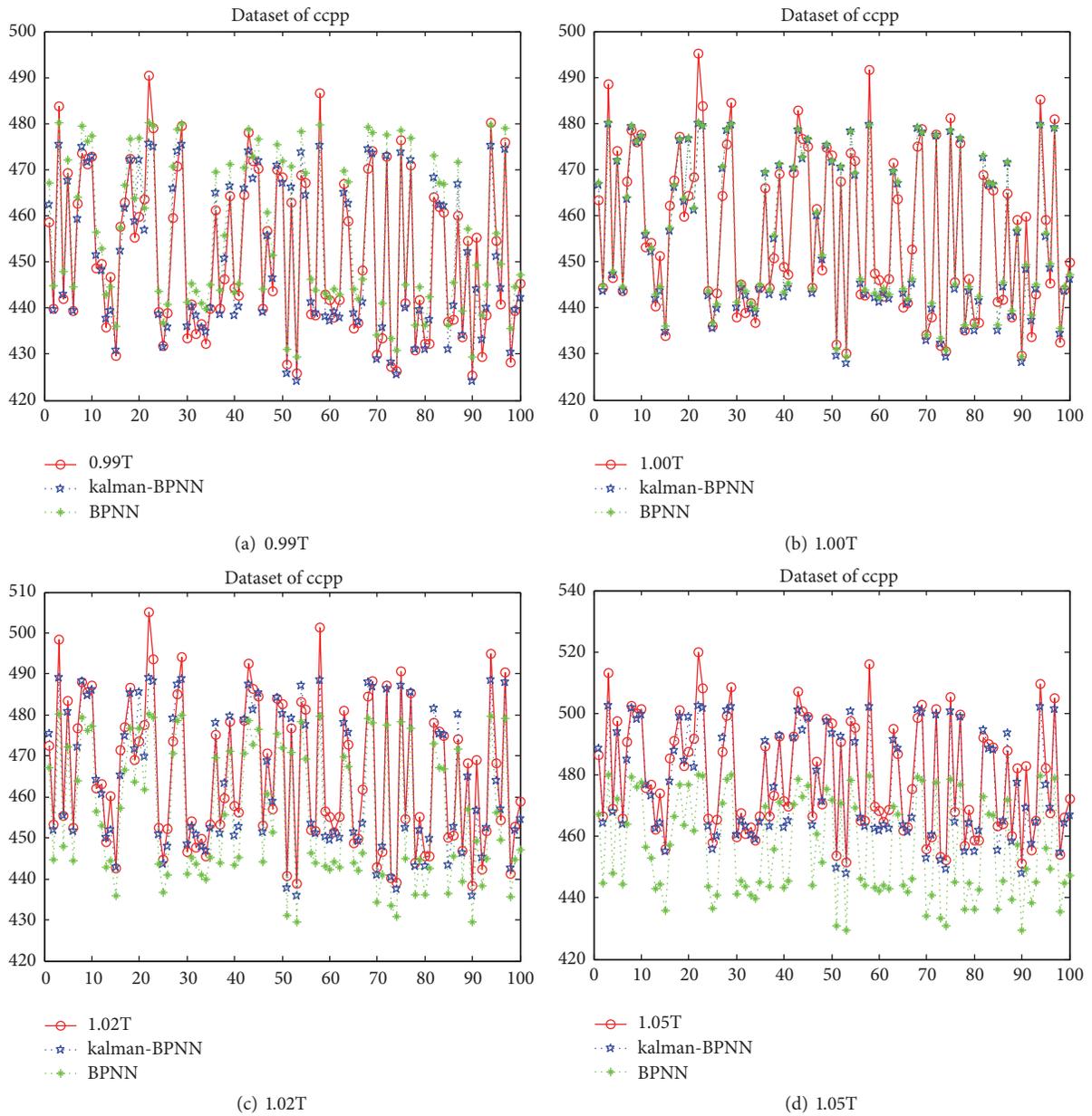


FIGURE 4: Neural network fitting output before and after the parameters of neural network were updated by Kalman filter after system modal change.

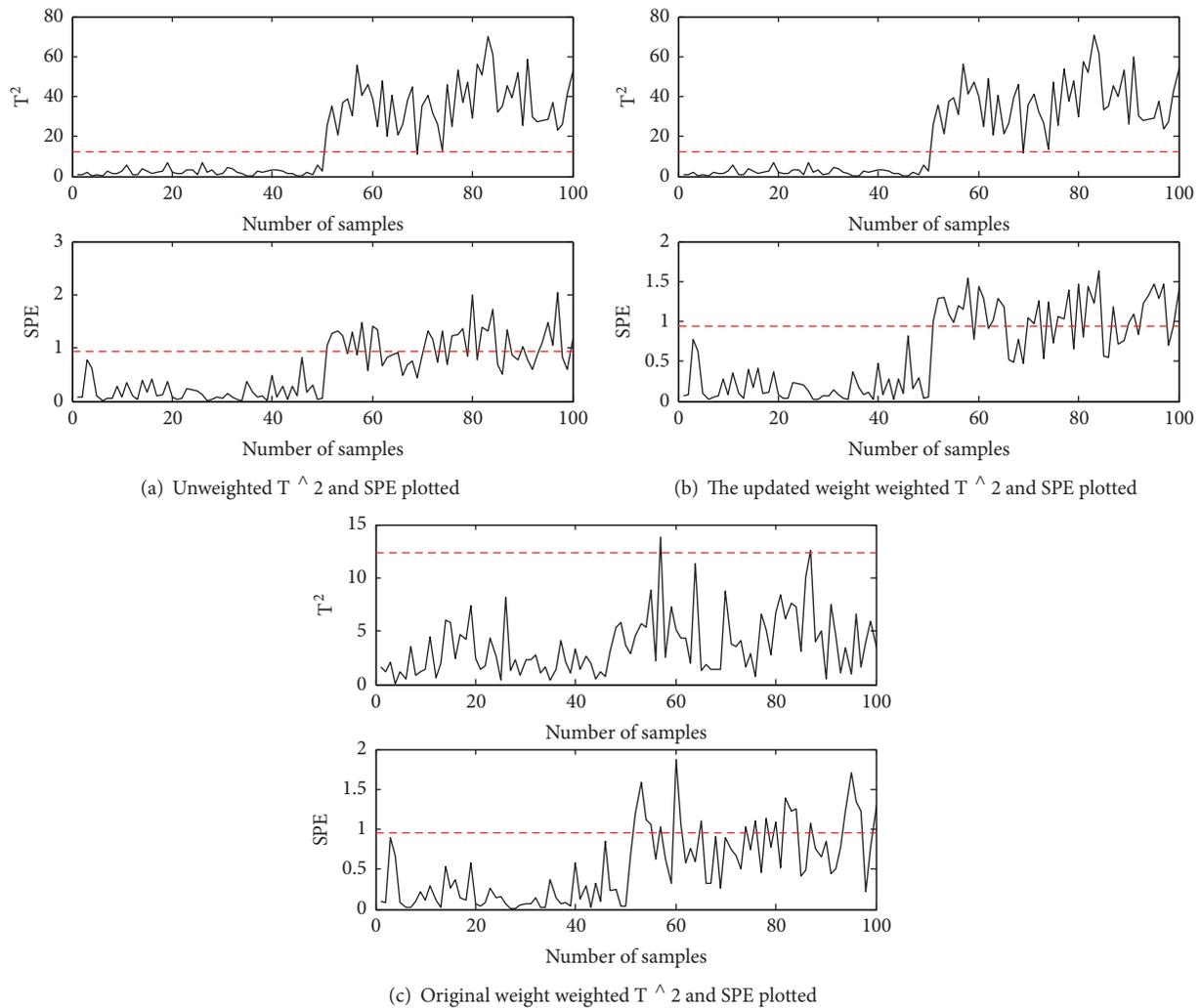


FIGURE 5: Fault detection diagram for 1.05T.

all parameters of the network using some nonlinear filtering algorithms [22].

6. Conclusions

This paper applies the Kalman filter to the real-time update of the hidden layer output weights of BP neural network, which greatly improves the applicability and real-time performance of the BP neural network and ensures the accuracy of the model parameters, so that the obtained evaluation function is accurate. This also makes it possible to characterize the impact of each feature variable on the output based on the weights of the data features obtained from the evaluation function. When there is not big change in the system state, the output of BP neural network hidden layer is adjusted by Kalman filter to compensate the output of the system, in order to avoid retraining the network once the system modality changes. Retraining will result in a waste of resources and may not be tolerated even with frequent retraining of some system models. The simulation results on a standard UCI-ccpp dataset validate the effectiveness of the proposed algorithm.

Once the training of neural network is completed, it is always a difficult problem to retrain global parameters and update local parameters after it is applied to practical industrial systems. The proposed algorithm only completes the real-time update of the hidden layer output weights of the BP neural network without changing other parameters in the network. On this basis, in the next step we are going to combine some nonlinear filtering algorithms and contents of incremental learning to seek a breakthrough in the regulation and updating of global network parameters.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work is supported by the National Natural Science Foundation (NNSF) of China under Grants nos. U1509203, 61490701, 61573137, 61673160, and 61333005.

References

- [1] S. Ferrari and R. F. Stengel, "Smooth function approximation using neural networks," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 16, no. 1, pp. 24–38, 2005.
- [2] T.-Y. Kwok and D.-Y. Yeung, "Objective functions for training new hidden units in constructive neural networks," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 8, no. 5, pp. 1131–1148, 1997.
- [3] S. Ganjefar and M. Tofghi, "Optimization of quantum-inspired neural network using memetic algorithm for function approximation and chaotic time series prediction," *Neurocomputing*, vol. 291, pp. 175–186, 2018.
- [4] Y. W. Wong, K. P. Seng, and L.-M. Ang, "Radial basis function neural network with incremental learning for face recognition," *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, vol. 41, no. 4, pp. 940–949, 2011.
- [5] L. Fu, "Incremental knowledge acquisition in supervised learning networks," *IEEE Transactions on Systems, Man and Cybernetics, Part A: Systems and Humans*, vol. 26, no. 6, pp. 801–809, 1996.
- [6] K. W. Kow, Y. W. Wong, R. K. Rajkumar, R. K. Rajkumar, and D. Isa, "Incremental unsupervised learning algorithm for power fluctuation event detection in PV grid-tied systems," *Lecture Notes in Electrical Engineering*, vol. 398, pp. 673–679, 2017.
- [7] H. Wen, W. Xie, J. Pei, and L. Guan, "An incremental learning algorithm for the hybrid RBF-BP network classifier," *EURASIP Journal on Advances in Signal Processing*, vol. 2016, no. 1, 2016.
- [8] S. Chen, C. F. N. Cowan, and P. M. Grant, "Orthogonal least squares learning algorithm for radial basis function networks," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 2, no. 3, pp. 302–309, 1991.
- [9] S. Wu and M. J. Er, "Dynamic fuzzy neural networks—a novel approach to function approximation," *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, vol. 30, no. 2, pp. 358–364, 2000.
- [10] S. Q. Wu, M. J. Er, and Y. Gao, "A fast approach for automatic generation of fuzzy rules by generalized dynamic fuzzy neural networks," *IEEE Transactions on Fuzzy Systems*, vol. 9, no. 4, pp. 578–594, 2001.
- [11] M. Zhang, X. Ma, B. Qin et al., "Information fusion control with time delay for smooth pursuit eye movement," *Physiological Reports*, vol. 4, no. 10, Article ID e12775, 2016.
- [12] L.-Z. Xu, X.-L. Feng, and C.-L. Wen, "Sequential fusion filtering for networked multi-sensor systems based on noise estimation," *Tien Tzu Hsueh Pao/Acta Electronica Sinica*, vol. 42, no. 1, pp. 160–168, 2014.
- [13] S. Ding and Q.-H. Wu, "A MATLAB-based study on approximation performances of improved algorithms of typical BP neural networks," *Applied Mechanics and Materials*, vol. 313–314, pp. 1353–1356, 2013.
- [14] G. Huang, L. Chen, and C. Siew, "Universal approximation using incremental constructive feedforward networks with random hidden nodes," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 17, no. 4, pp. 879–892, 2006.
- [15] S. Rhode, K. Usevich, I. Markovsky, and F. Gauterin, "A recursive restricted total least-squares algorithm," *IEEE Transactions on Signal Processing*, vol. 62, no. 21, pp. 5652–5662, 2014.
- [16] W. Chenglin, *Multi-Scale Dynamic Modeling Theory and Its Applications*, Science Press, 2008.
- [17] Q. Liu, Z. Wang, X. He, and D. Zhou, "On Kalman-consensus filtering with random link failures over sensor networks," *IEEE Transactions on Automatic Control*, vol. 99, pp. 1–1, 2017.
- [18] Q. Liu, Z. Wang, X. He, G. Ghinea, and F. E. Alsaadi, "A resilient approach to distributed filter design for time-varying systems under stochastic nonlinearities and sensor degradation," *IEEE Transactions on Signal Processing*, vol. 65, no. 5, pp. 1300–1309, 2017.
- [19] C. Wen, Z. Wang, J. Hu, Q. Liu, and F. E. Alsaadi, "Recursive filtering for state-saturated systems with randomly occurring nonlinearities and missing measurements," *International Journal of Robust and Nonlinear Control*, vol. 28, no. 5, pp. 1715–1727, 2018.
- [20] S. Ji, X. Xu, and C. Wen, "A kind of K—Nearest neighbor fault diagnosis method based on MIV data transformation," in *Proceedings of the 2017 Chinese Automation Congress (CAC)*, pp. 6306–6310, Jinan, October 2017.
- [21] M. Lichman, *UCI Machine Learning Repository*, University of California, School of Information and Computer Science, Irvine, Calif, USA, 2013, <http://archive.ics.uci.edu/ml>.
- [22] C. Wen, Z. Wang, T. Geng, and F. E. Alsaadi, "Event-based distributed recursive filtering for state-saturated systems with redundant channels," *Information Fusion*, vol. 39, pp. 96–107, 2018.

Research Article

Fault Diagnosis Method Based on Gap Metric Data Preprocessing and Principal Component Analysis

Zihan Wang, Chenglin Wen , Xiaoming Xu , and Siyu Ji 

School of Automation, Hangzhou Dianzi University, Hangzhou 310018, China

Correspondence should be addressed to Chenglin Wen; wenc1@hdu.edu.cn

Received 22 February 2018; Accepted 11 April 2018; Published 17 May 2018

Academic Editor: Zhijie Zhou

Copyright © 2018 Zihan Wang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Principal component analysis (PCA) is widely used in fault diagnosis. Because the traditional data preprocessing method ignores the correlation between different variables in the system, the feature extraction is not accurate. In order to solve it, this paper proposes a kind of data preprocessing method based on the Gap metric to improve the performance of PCA in fault diagnosis. For different types of faults, the original dataset transformation through Gap metric can reflect the correlation of different variables of the system in high-dimensional space, so as to model more accurately. Finally, the feasibility and effectiveness of the proposed method are verified through simulation.

1. Introduction

As the complexity of industrial manufacturing systems increases, the correlation between variables in the system becomes more complex, and these variables contain important information about the status of the system. Therefore, it is an important issue that fault detection and diagnosis of the system are through the information of these variables.

However, in industrial manufacturing systems, because of the different dimensions of system variables, it is usually necessary to preprocess the data to standardize the data. In traditional data preprocessing methods, ignoring the influence of dimension on the correlation between system variables leads to the lack of correlation of system variables after data preprocessing, which makes it difficult to extract the representative principal components. Therefore, maintaining the correlation between system variables is the key to data preprocessing.

In order to solve this problem, many studies have been made. Wen et al. proposed a method called Relative Principle Component Analysis (RPCA) [1]; it introduces analyzing and determining the importance of each component to the prior information of the system, giving the corresponding weight of each component of the system, and establishing relative

principal component model. Literature [2] proposed a fault diagnosis method based on information incremental matrix. Based on this, Yuan et al. proposed a relative transformation of information incremental matrix fault diagnosis method [3], which can effectively detect variables that play an important role in the system. Because of their smaller absolute value and less absolute changes, small changes of these important variables usually play a very crucial role in the system. Xu and Wen proposed a fault diagnosis method based on information entropy and relative principal component analysis [4]; in high-dimensional system, the high correlation of system variables leads to the model not being able to select the representative principal components. The approach given by them is to use information entropy to measure the uncertainty of variables and calculate the information gain of variables. According to the different degrees of importance of variables, relatively transform the data to get a more accurate data model. Jiao et al. proposed a method for simulation model validation based on Theil's inequality coefficient and principal component analysis [5]; it is based on the TIC model; given a model of the differences in position and trend between the simulated output and the reference output, there is a correlation between the two differences, using PCA to obtain the verification results. Kangling et al. proposed a

fault diagnosis study based on adaptive partition PCA [6]. In order to solve the inaccurate modeling problem, the diagnosis model can be automatically updated and adjusted, so as to improve the model matching and the accuracy of diagnosis results.

The Gap metric is proved to be more suitable for measuring the distance between two linear systems than the norm-based ones [7, 8], and the effect of dimension on each variable can be reflected in the Riemannian space when data preprocessing is performed. The gap metric is widely used in the study of the uncertainty and robustness of the feedback system. Tryphon proposed a method that can easily calculate the gap metric [9], which improves the practicality of the gap metric. Literature [10] proposed the concept of ν -gap metric in the frequency domain and then extended it to nonlinear systems [11]. Ebadollahi and Saki proposed that in the multimodel predictive control, in order to track the maximum power point without losing its control performance, the gap metric method is used to divide the entire area of the partial load system into a corresponding linear model. This method ensures the stability of the original closed-loop system [12]. Konghuayrob and Kaitwanidvilai used ν -gap to measure the distance between two linear systems [13]. They used low-order controllers with similar dynamic characteristics to replace traditional high-level and complex controllers, and both controllers have similar dynamic characteristics and robustness. For multilinear model control of nonlinear systems, Du proposed a weighting method based on $1/\delta$ Gap metric [14]. The Gap metric method was used to calculate the weighting function of the local controller combination. The validity of the method was verified by the CSTR system.

In the principal component analysis method based on the traditional data preprocessing, when removing the original data dimension, all are based on the European metric data preprocessing method, and some important information of the data will be ignored, and often these data are important variables that contain slowly changing fault information. In the method proposed in this article, Gap metric can project data on Riemann spheres in Riemann space and can highlight the information easily ignored in the European space. After the Gap metric data preprocessing method is adopted, the eigenvalue feature vector decomposition is performed on the processed data matrix, and the principal component is constructed to construct the principal component space according to the cumulative percent variance criterion, since Gap metric can highlight the variable data with small absolute change and relatively large variation in the system variables, so we can extract the main component and the main component vector when constructing the principal component space. By calculating the T^2 statistical limit and the SPE statistical limit of the normal system dataset, we detect whether the T^2 statistics and SPE statistics of the test dataset exceed the limit to judge whether the system is faulty, and then the fault variables are separated by the contribution of system variables to fault samples.

The rest of this article is organized as follows. In the second section, we briefly review the PCA approach. In the third section, we propose a kind of improved PCA data

preprocessing method, which is data preprocessing method based on gap metric. In the fourth section, we set up a system model and test the feasibility and effectiveness of the proposed method by different types of faults. In the fifth section, we give a summary and future research direction.

2. PCA Based on Traditional Data Preprocessing

The basic idea of PCA is to decompose multivariable sample space into lower dimensional principal component subspaces composed of principal components variables and a residual subspace according to the historical data of process variables. And the statistics which can reflect the change of space are constructed in these two subspaces. Then, the sample vectors are, respectively, projected into two subspaces and compute the distance from the sample to the subspace. The process monitoring and fault detection are performed by comparing the distance with the corresponding statistics.

First, model by PCA to variable space. Select a set of variables under normal conditions as the original data. $x \in R^m$ is a test sample that contains m variables and each variable has n independent samples. Construct the original measurement data matrix:

$$\mathbf{X}_n = [x(1), x(2), \dots, x(n)], \quad \mathbf{X}_n \in R^{m \times n}. \quad (1)$$

Here, each column of \mathbf{X}_n represents a variable, and each row represents a sample. Because the dimensions of the measured variables are different, each column of the data matrix is normalized. Assuming that the normalized measurement data matrix is \mathbf{X}^*

$$\mathbf{X}^* = \frac{\mathbf{X}_n - e \cdot q}{\text{diag}(\sigma_1, \sigma_2, \dots, \sigma_n)}. \quad (2)$$

Here, $e = (1, 1, \dots, 1)^T \in R^{m \times 1}$, $q = (q_1, q_2, \dots, q_n) \in R^{1 \times n}$, q is the average of X columns, $i = 1, \dots, n$. $\text{diag}(\sigma_1, \sigma_2, \dots, \sigma_n)$ is the variance matrix of X_n .

The covariance matrix \mathbf{S} of \mathbf{X}^* is

$$\mathbf{S} = \frac{1}{n-1} \mathbf{X}^{*T} \mathbf{X}^*. \quad (3)$$

The processing of the matrix \mathbf{S} is generally eigenvalue decomposition, according to the size of the eigenvalues arranged in descending order. The PCA model decomposes \mathbf{X}^* as follows:

$$\begin{aligned} \mathbf{X}^* &= \widehat{\mathbf{X}}^* + \mathbf{E} = \mathbf{T} \mathbf{P}^T + \mathbf{E}, \\ \mathbf{T} &= \mathbf{X}^* \mathbf{P}, \end{aligned} \quad (4)$$

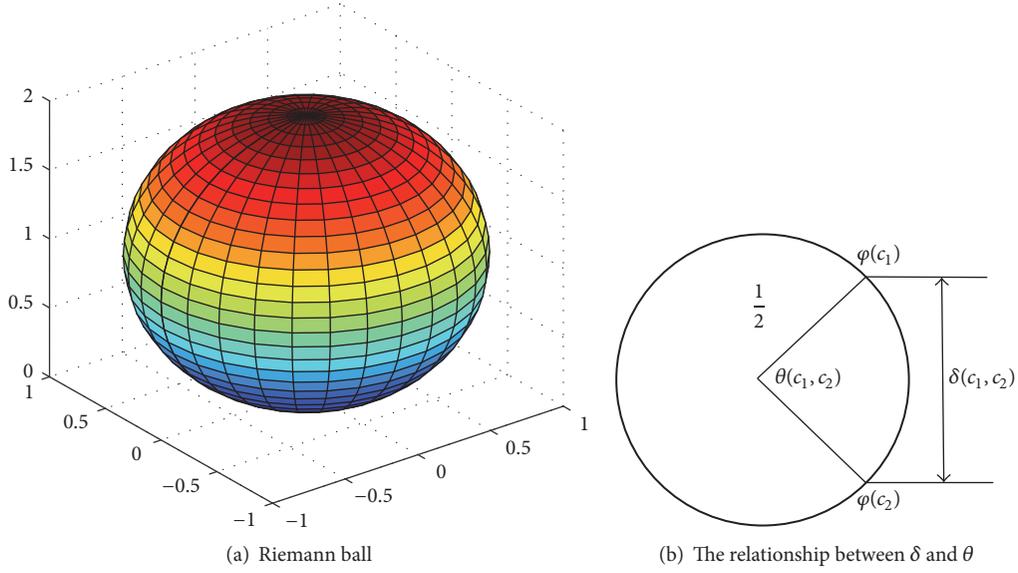


FIGURE 1: The geometric meaning of the gap metric.

where $\widehat{\mathbf{X}}^*$ is the projection in the principal component space, \mathbf{E} is the projection in the residual space, and $\mathbf{P} \in \mathbf{R}^{m \times A}$ is the load matrix, which consists of the first A eigenvectors of \mathbf{S} . $\mathbf{T} \in \mathbf{R}^{m \times A}$ is the scoring matrix, the elements of \mathbf{T} are called the primary variables, and A is the number of the primary. The principal component space is the part of the modeling, and the residual space is not the part of the modeling, which represents the noise and fault information in the data.

The selection of the principal component number A is based on the Cumulative Percent Variance (CPV). This criterion determines the number of principal elements based on the cumulative sum of percentages of principal components. The CPV represents the ratio of the data changes explained by the first A principal component to the total data changes. Therefore, the cumulative contribution rate CPV of the first A principal can be expressed as

$$\text{CPV} = \frac{\sum_{i=1}^A \lambda_i}{\sum_{i=1}^m \lambda_i}, \quad (5)$$

where λ_i is the eigenvalue of the covariance matrix \mathbf{S} . In general, when the cumulative contribution rate reaches 85% or more, it is considered that the number of elements A contains enough information of the original data.

3. Improved Data Preprocessing Method

In this section, we propose a kind of data preprocessing method based on Gap metric.

3.1. Gap Metric. In the Riemann space, $\varphi(c_1)$ and $\varphi(c_2)$ are used to represent the spherical projection of complex numbers c_1 and c_2 on a three-dimensional Riemann ball with

diameter 1 and the chord between c_1 and c_2 is denoted by $\delta(c_1, c_2)$; then $\delta(c_1, c_2)$ is defined by

$$\delta(c_1, c_2) = \|\varphi(c_1) - \varphi(c_2)\| = \frac{|c_1 - c_2|}{\sqrt{1 + c_1^2} \sqrt{1 + c_2^2}}. \quad (6)$$

$\theta(c_1, c_2)$ is used to express the spherical distance between c_1 and c_2 , that is, the arc length connecting $\varphi(c_1)$ and $\varphi(c_2)$ on the Riemann ball; then

$$\theta(c_1, c_2) = \arcsin \delta(c_1, c_2) = \arcsin \frac{|c_1 - c_2|}{\sqrt{1 + c_1^2} \sqrt{1 + c_2^2}}. \quad (7)$$

As can be seen from Figure 1, the shortest arc length on the circle is obtained from a plane-cut Riemann sphere determined by 3 points of the center of the ball, $\varphi(c_1)$ and $\varphi(c_2)$.

The character of the Gap metric in the control system has similar properties in the data space. The nature of the Gap metric in the data space is as follows:

- (1) Gap metric can be regarded as the distance characterization of data in Riemann space, which is an extension to the traditional method based on infinite norm metrics.
- (2) The value of the Gap metric is in the range of 0 to 1. The smaller the value, the closer the characteristics of the two datasets. The larger the value, the greater the difference in the characteristics of the two datasets. If the Gap metric of two datasets is 0, then they contain exactly the same characteristics.

3.2. *Gap Metric Data Preprocessing and Fault Diagnosis.* Set the data observation matrix $X_n \in R^{m \times n}$ of the multivariable system as

$$\mathbf{X}_n = \begin{bmatrix} x_1(1) & x_1(2) & \cdots & x_1(n) \\ x_2(1) & x_2(2) & \cdots & x_2(n) \\ \vdots & \vdots & \ddots & \vdots \\ x_m(1) & x_m(2) & \cdots & x_m(n) \end{bmatrix}. \quad (8)$$

Here the column vector $x_i(j) = [x_1(j), x_2(j), \dots, x_m(j)]^T$, $j = 1, 2, \dots, n$, represents the system variable, and the row

vector represents sampled data at a sampling instant. Then, preprocessing the data matrix, the mean vector of \mathbf{X}_n is

$$\mathbf{b}_n = \frac{1}{m} \mathbf{1}_m \mathbf{X}_n. \quad (9)$$

where $\mathbf{1}_m = [1, 1, \dots, 1] \in R^{1 \times m}$.

Step 1. Project the original data onto the Riemann sphere and calculate the gap metric for each variable, and the matrix \mathbf{X}^* is calculated:

$$\mathbf{X}^* = \begin{bmatrix} \delta(x_1(1), \mathbf{b}_n(1)) & \delta(x_1(2), \mathbf{b}_n(2)) & \cdots & \delta(x_1(n), \mathbf{b}_n(n)) \\ \delta(x_2(1), \mathbf{b}_n(1)) & \delta(x_2(2), \mathbf{b}_n(2)) & \cdots & \delta(x_2(n), \mathbf{b}_n(n)) \\ \vdots & \vdots & \ddots & \vdots \\ \delta(x_m(1), \mathbf{b}_n(1)) & \delta(x_m(2), \mathbf{b}_n(2)) & \cdots & \delta(x_m(n), \mathbf{b}_n(n)) \end{bmatrix}. \quad (10)$$

Here

$$\begin{aligned} \delta(x_i(k), \mathbf{b}_n(k)) &= \|\varphi(x_i(k)) - \varphi(\mathbf{b}_n(k))\| \\ &= \frac{|x_i(k) - b_n(k)|}{\sqrt{1 + (x_i(k))^2} \cdot \sqrt{1 + (b_n(k))^2}}. \end{aligned} \quad (11)$$

Step 2. Decompose eigenvalues and eigenvectors of \mathbf{X}^* , select pivotal elements based on Cumulative Percent Variance (CPV), and construct principal component space and residual space.

Step 3. Calculate the statistical limit of T^2 based on the principal component space of normal data and calculate the *SPE* statistical limit through the residual space.

Step 4. Set the test data matrix $\mathbf{Y}_n \in R^{m \times n}$ as

$$\mathbf{Y}_n = \begin{bmatrix} y_1(1) & y_1(2) & \cdots & y_1(n) \\ y_2(1) & y_2(2) & \cdots & y_2(n) \\ \vdots & \vdots & \ddots & \vdots \\ y_m(1) & y_m(2) & \cdots & y_m(n) \end{bmatrix}. \quad (12)$$

Preprocess the test dataset \mathbf{Y}_n with Gap metric method to get \mathbf{Y}^* :

$$\mathbf{Y}^* = \begin{bmatrix} \delta(y_1(1), \mathbf{b}_n(1)) & \delta(y_1(2), \mathbf{b}_n(2)) & \cdots & \delta(y_1(n), \mathbf{b}_n(n)) \\ \delta(y_2(1), \mathbf{b}_n(1)) & \delta(y_2(2), \mathbf{b}_n(2)) & \cdots & \delta(y_2(n), \mathbf{b}_n(n)) \\ \vdots & \vdots & \ddots & \vdots \\ \delta(y_m(1), \mathbf{b}_n(1)) & \delta(y_m(2), \mathbf{b}_n(2)) & \cdots & \delta(y_m(n), \mathbf{b}_n(n)) \end{bmatrix}. \quad (13)$$

Here, \mathbf{b}_n is the mean vector of \mathbf{X}_n .

Step 5. Calculate the T^2 statistic and *SPE* statistic separately and detect whether the statistic exceeds the statistical limit of normal data. If the statistic exceeds the statistical limit, it will be faulty. Otherwise, it will be normal.

Step 6. Fault variables can be separated by the contribution of system variables to the fault in the residual space.

The physical meaning of Gap metric is the chord pitch of the data projected on the Riemann sphere through a spherical surface, which can highlight the impact of the

TABLE 1: Detection result.

| Constant deviation error | PCA | | InEnPCA | | GAPPCA | |
|---------------------------|-------|-------|---------|-------|--------|-------|
| | T^2 | SPE | T^2 | SPE | T^2 | SPE |
| Misdiagnosis | 8 | 7 | 7 | 8 | 16 | 8 |
| Rate of misdiagnosis | 0.010 | 0.009 | 0.009 | 0.010 | 0.020 | 0.010 |
| Omissive judgement | 54 | 0 | 66 | 0 | 8 | 0 |
| Rate of omissive judgment | 0.270 | 0 | 0.330 | 0 | 0.040 | 0 |

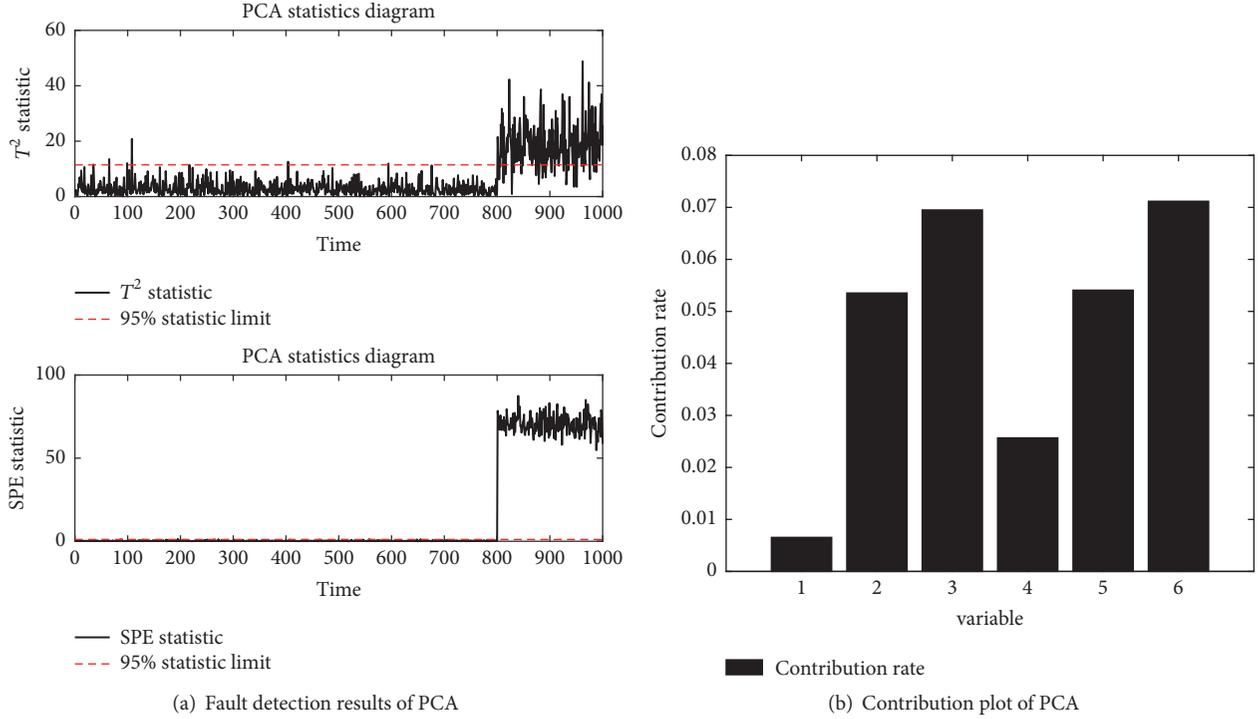


FIGURE 2: PCA with constant deviation fault in 801~1000 samples.

relative changes of the data on its own. The transformed data does not ignore variables that have small absolute changes but relatively large transformations. Frequently, these variables also contain important information.

4. Simulation

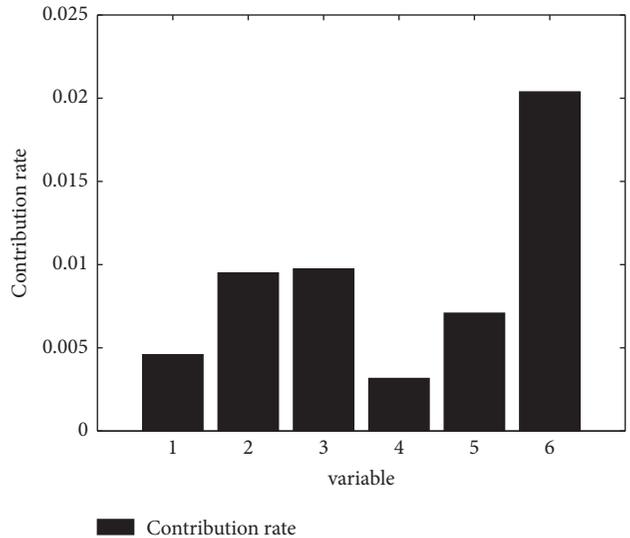
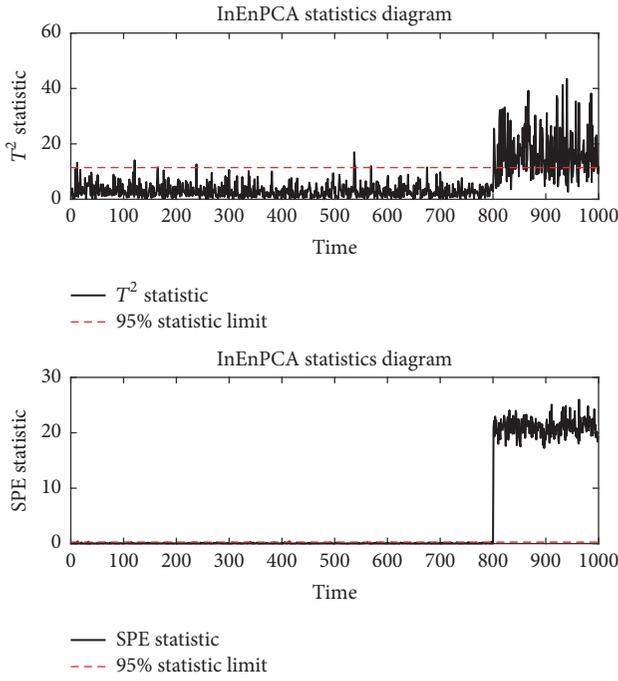
In order to verify the effectiveness of the proposed method, considering random variables and their linear combinations, 6 system variables are constructed as follows:

$$\begin{aligned}
 \mathbf{x}_1 &= 0.1 \times \text{randn}(1, n), \\
 \mathbf{x}_2 &= 0.2 \times \text{randn}(1, n), \\
 \mathbf{x}_3 &= 0.3 \times \text{randn}(1, n), \\
 \mathbf{x}_4 &= -1.3\mathbf{x}_1 + 0.2\mathbf{x}_2 + 0.8\mathbf{x}_3 + 0.1 \times \text{randn}(1, n), \\
 \mathbf{x}_5 &= \mathbf{x}_2 - 0.3\mathbf{x}_3 + 0.1 \times \text{randn}(1, n), \\
 \mathbf{x}_6 &= \mathbf{x}_1 + \mathbf{x}_4 + 0.1 \times \text{randn}(1, n).
 \end{aligned} \tag{14}$$

Among them, $\text{randn}(1, n)$ is a random sequence of 1 row and N columns generated by MATLAB. First, choose 1000 normal samples to establish the PCA model, then select 1000 samples as the test data, and introduce the constant deviation error of magnitude 3 to the last 200 samples of the variable \mathbf{x}_6 in the test data, and detect them, respectively, with the model. SPE statistics and Hotelling's T^2 statistics are used as indicators to measure the number of misinformation and the missing number of PCA.

As shown in Table 1, the average misinformation and the number of misstatements of the PCA, InEnPCA, and GAPPCA methods are obtained through 10 simulation statistics.

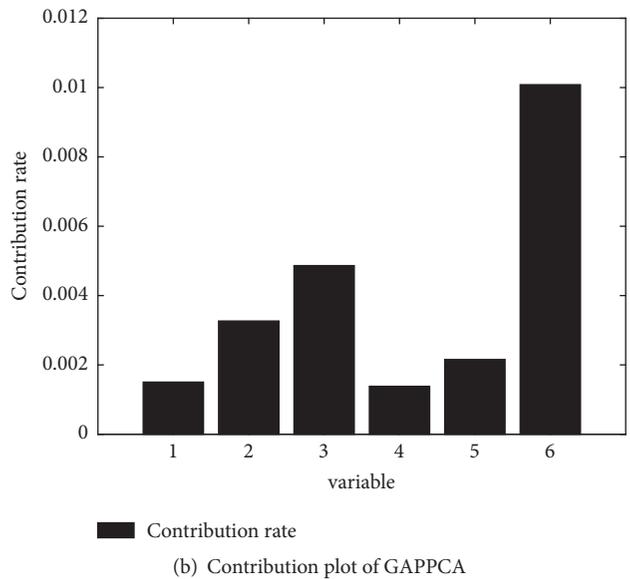
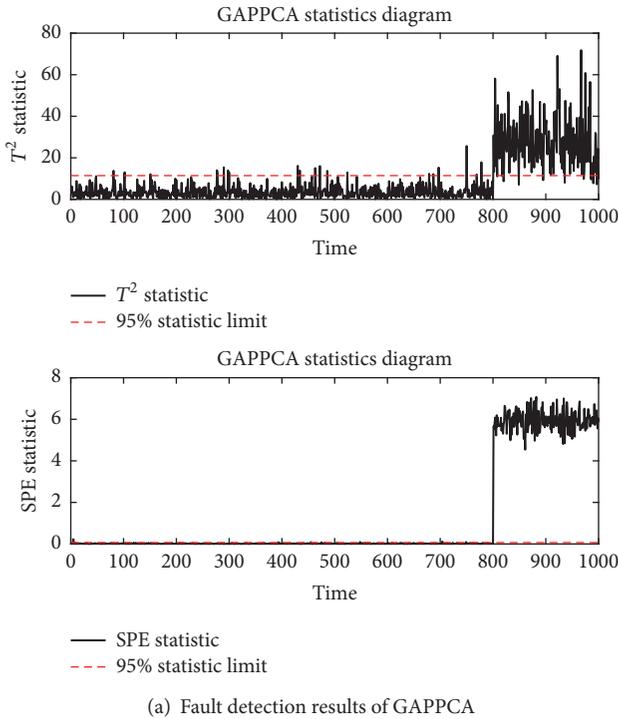
In the test, 1~800 times exceeds the control limit for misdiagnosis, 801~1000 times below the control limit for omissive judgement. It can be seen from Figures 2~4 that the misdiagnosis rate of PCA, InEnPCA, and GAPPCA is not higher than 2%. However, in the fault omission detection, T^2 statistics of traditional PCA and InEnPCA show a large number of fault omissions. The omissive judgement rate of T^2 in PCA and InEnPCA is 27% and 33%, respectively, while that of GAPPCA is only 4%; this shows that after the



(a) Fault detection results of InEnPCA

(b) Contribution plot of InEnPCA

FIGURE 3: InEnPCA with constant deviation fault in 801~1000 samples.



(a) Fault detection results of GAPPCA

(b) Contribution plot of GAPPCA

FIGURE 4: GAPPCA with constant deviation fault in 801~1000 samples.

data is preprocessed by gap metric, the principal component model established contains most of the key information of the system, thus improving the accuracy of fault detection. In the fault diagnosis, it can be seen from the contribution graph of

the fault that the contribution rate of the fault variable x in GAPPCA is more easily distinguished from the contribution of other variables than PCA and InEnPCA; because system variables are preprocessed for gap metric projection on a

TABLE 2: Detection result.

| Constant deviation error | PCA | | InEnPCA | | GAPPCA | |
|---------------------------|-------|-------|---------|-------|--------|-------|
| | T^2 | SPE | T^2 | SPE | T^2 | SPE |
| Misdiagnosis | 12 | 9 | 8 | 8 | 12 | 13 |
| Rate of misdiagnosis | 0.015 | 0.011 | 0.010 | 0.010 | 0.015 | 0.016 |
| Omissive judgement | 199 | 81 | 134 | 62 | 75 | 0 |
| Rate of omissive judgment | 0.995 | 0.405 | 0.670 | 0.310 | 0.094 | 0 |

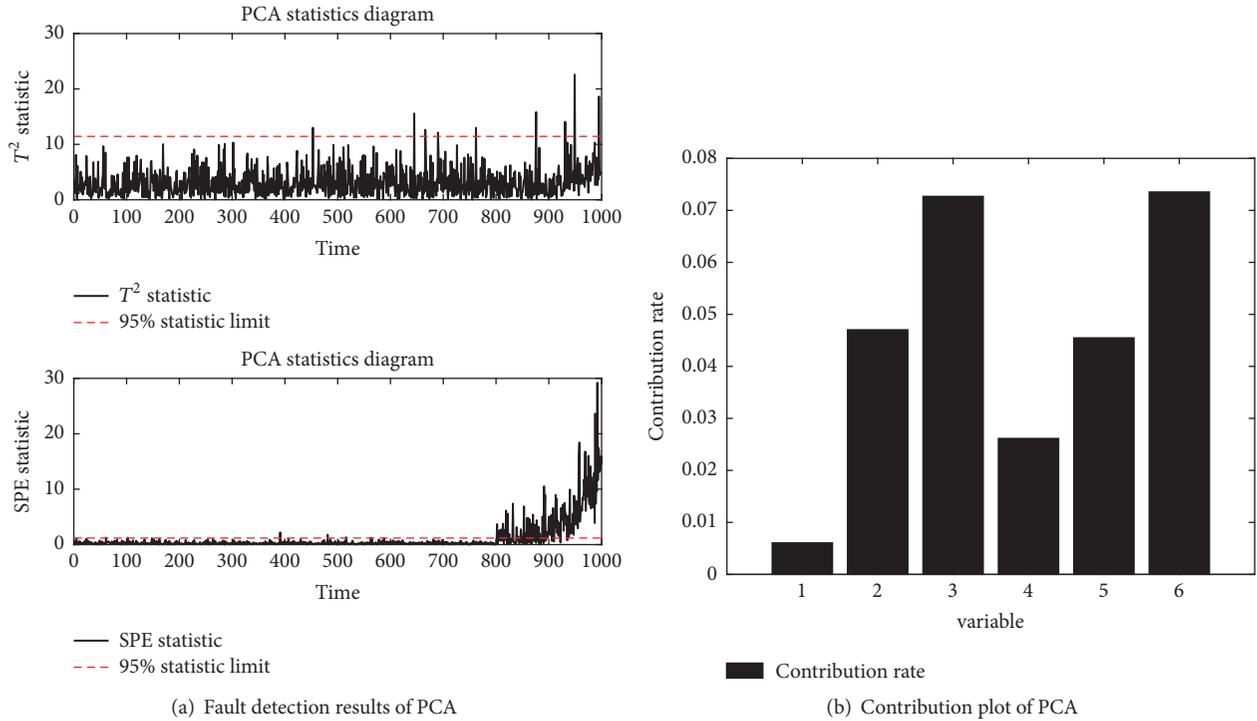


FIGURE 5: PCA with microfaults detection in 801~1000 samples.

Riemann sphere, the correlation between system variables can be better reflected.

The application of traditional PCA method in fault diagnosis is based on the absolute distance between samples as the criterion for fault detection and fault diagnosis. However, in real systems, the occurrence of microfaults often changes very little. Therefore, the detection of minor faults is extremely necessary. Literature [15] combines the PCA technique with the univariate exponential weighted-sliding averaging for the correlation of variables in the chemical process. In view of the small deviation between normal and microfaults, [16] combines probability distribution metric and Kullback-Leibler measure to quantify residuals between potential scores and reference scores and proposed PCA algorithm control limits for small faults. Literature [17] is an analysis model based on literature [16]. The authors [17] give an approach where Kullback-Leibler divergence (KLD) was applied to the principal component variables obtained after process dimensionality reduction using PCA. This method was used to diagnose microfaults.

In order to verify the detection accuracy in the case of a microfault, a slight fault with a slowly increasing rate of 0.1% is introduced into the last 200 sample points of the variable x_6 in the test data. Use this model to test the fault diagnosis performance of the three methods.

As shown in Table 2, the average misinformation and the number of misstatements of the PCA, InEnPCA, and GAPPCA methods are obtained through 10 simulation statistics.

It can be seen from Figures 5–7 that in the detection of microfaults, due to the small changes in system variables, the traditional PCA method cannot extract the representative of the main element; as can be seen from Figure 6, PCA cannot reflect the system fault status during fault detection and diagnosis. Because the relative change of microfaults is bigger than absolute change, GAPPCA can extract representative principal component variables for small changes of system variables, and the detection results are better than the traditional PCA. In fault diagnosis, InEnPCA reflects the information gain of system variables and thus has good performance in fault diagnosis. Besides, since the gap metric

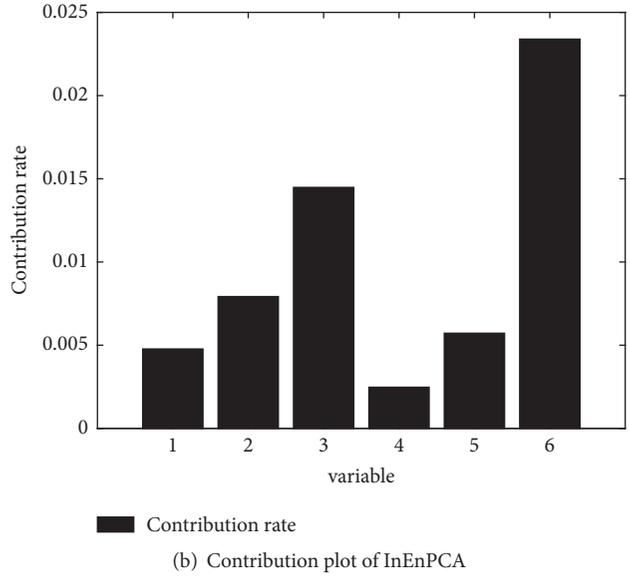
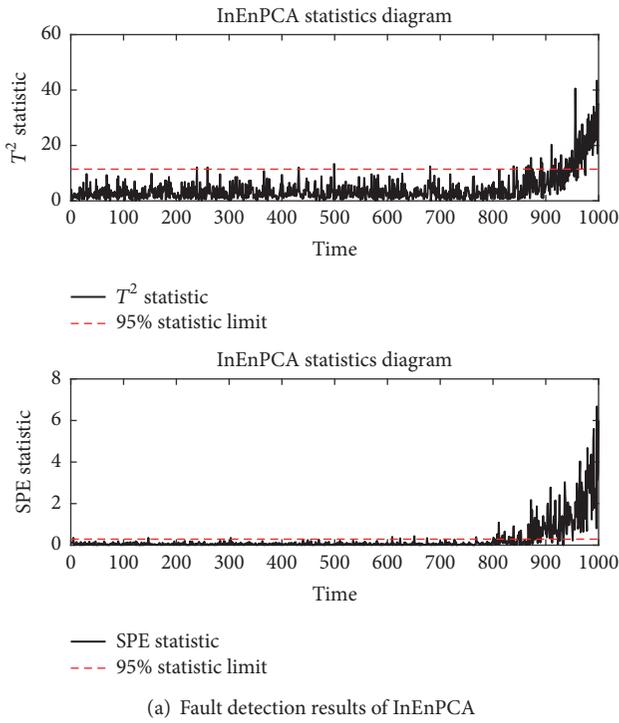


FIGURE 6: InEnPCA with microfaults detection in 801~1000 samples.

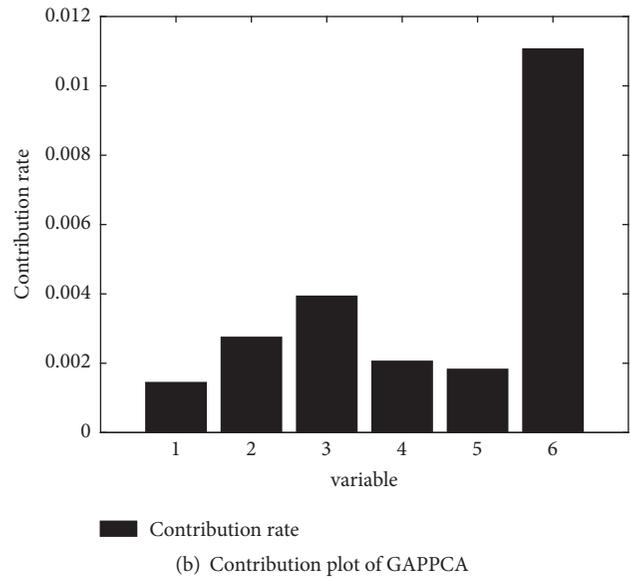
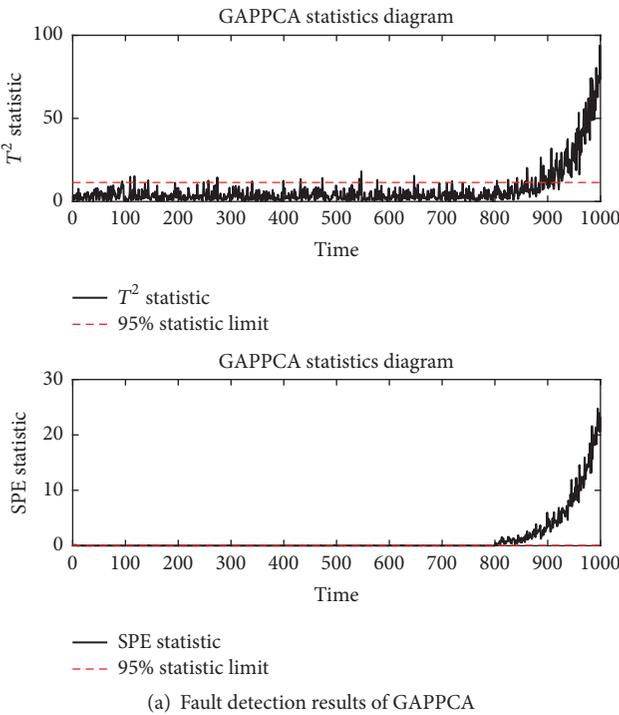


FIGURE 7: GAPPCA with microfaults detection in 801~1000 samples.

reflects the correlation of system variables in the Riemannian space better than those in the Euclidean space, when using the contribution graph to separate the fault variables, we can see that the contribution rate of fault variable x_6 in the contribution graph of GAPPCA is remarkable.

To sum up, compared with the traditional PCA preprocessing method in Euclidean space, the preprocessing method based on the gap metric can better reflect the relevant information between the variables, and the faults that occur to some small but important variables can be diagnosed accurately.

5. Conclusions

In this paper, we propose a fault diagnosis method based on Gap metric data preprocessing and PCA. When some variables in the system play an important role, and the absolute value of the variables themselves is small, they cannot detect the smaller faults. The proposed method can detect the source of the fault more accurately and reduce the rate of misdiagnosis and rate of ommissive judgement.

However, there are still some problems to be studied in this paper. In the PCA fault diagnosis method based on the Gap metric data preprocessing, there will be a problem that the projection overlaps. The fault information reflected by the fault samples may be projected after the Gap metric processing becomes normal. Within the region, how to separate fault information and normal information in high-dimensional space is the focus of further research.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

This work is supported by the National Natural Science Foundation (NNSF) of China under Grants nos. U1509203, 61333005, 61490701, 61573137, and 61673160.

References

- [1] C. L. Wen, J. Hu, T. Z. Wang, and Z. Chen, "Relative PCA with applications of data compression and fault diagnosis," *Acta Automatica Sinica*, vol. 34, no. 9, pp. 1128–1139, 2008.
- [2] C. L. Wen and Y. C. Hu, "Fault diagnosis based on the information incremental matrix," *Training of research workers in the medical sciences*. World Health Organization, vol. 38, no. 5, pp. 832–840, 2012.
- [3] T. Yuan, J. Hu, and C. Wen, "Fault diagnosis based on relative transformation of information increment matrix," *Journal of Central South University*, vol. 44, no. 1, pp. 238–242, 2013.
- [4] X. Xu and C. Wen, "Fault Diagnosis Method Based on Information Entropy and Relative Principal Component Analysis," *Journal of Control Science and Engineering*, pp. 1–8, 2017.
- [5] S. Jiao, W. Li, and M. Yang, "A Method for Simulation Model Validation Based on Theils Inequality Coefficient and Principal Component Analysis," *Communications in Computer Information Science*, pp. 402–126, 2013.
- [6] Kangling, Zhengshun, and J. Liang, "Adaptive partitioning PCA model for improving fault detection and isolation," *Chinese Journal of Chemical Engineering*, vol. 23, no. 6, article no. 235, pp. 981–991, 2015.
- [7] A. K. El-Sakkary, "The gap metric: robustness of stabilization of feedback systems," *Institute of Electrical and Electronics Engineers Transactions on Automatic Control*, vol. 30, no. 3, pp. 240–247, 1985.
- [8] T. T. Georgiou and M. C. Smith, "Optimal robustness in the gap metric," *Institute of Electrical and Electronics Engineers Transactions on Automatic Control*, vol. 35, no. 6, pp. 673–686, 1990.
- [9] T. T. Georgiou, "On the computation of the gap metric," in *Proceedings of the 27th IEEE Conference on Decision and Control*, pp. 1360–1361, December 1988.
- [10] G. Vinnicombe, "Frequency Domain Uncertainty and the Graph Topology," *IEEE Transactions on Automatic Control*, vol. 38, no. 9, pp. 1371–1383, 1993.
- [11] G. Vinnicombe, "v-gap distance for uncertain and nonlinear systems," in *Proceedings of the The 38th IEEE Conference on Decision and Control (CDC)*, pp. 2557–2562, December 1999.
- [12] S. Ebadollahi and S. Saki, "Wind Turbine Torque Oscillation Reduction Using Soft Switching Multiple Model Predictive Control Based on the Gap Metric and Kalman Filter Estimator," *IEEE Transactions on Industrial Electronics*, vol. 99, pp. 1–1, 2017.
- [13] P. Konghuayrob and S. Kaitwanidvilai, "Low order robust v-gap metric H ∞ loop shaping controller synthesis based on particle swarm optimization," *International Journal of Innovative Computing, Information and Control*, vol. 13, no. 6, pp. 1777–1789, 2017.
- [14] J. Du, "Comparison of two gap-metric-based weighting methods in multilinear model control," in *Proceedings of the 36th Chinese Control Conference, CCC 2017*, pp. 834–837, China, July 2017.
- [15] E. G. Zhi-Qiang, J. C. Yang, and Z. H. Song, "Research and Application of Small Shifts Detection Method Based on MEWMA-PCA," *Information & Control*, vol. 35, no. 9, pp. 650–656, 2007.
- [16] J. Harmouche, C. Delpha, and D. Diallo, "Incipient fault detection and diagnosis based on Kullback-Leibler divergence using Principal Component Analysis—part I," *Signal Processing*, vol. 94, no. 1, pp. 278–287, 2014.
- [17] J. Harmouche, C. Delpha, and D. Diallo, "Incipient fault detection and diagnosis based on Kullback-Leibler divergence using principal component analysis: Part II," *Signal Processing*, vol. 109, pp. 334–344, 2015.

Research Article

A New Residual Life Prediction Method for Complex Systems Based on Wiener Process and Evidential Reasoning

Xin Zhang,¹ Shuaiwen Tang,¹ Taoyuan Liu,¹ and Bangcheng Zhang ²

¹Rocket Force University of Engineering, Xi'an 710000, China

²School of Mechatronic Engineering, Changchun University of Technology, Changchun 130012, China

Correspondence should be addressed to Bangcheng Zhang; zhangbangcheng@ccut.edu.cn

Received 21 February 2018; Accepted 26 March 2018; Published 3 May 2018

Academic Editor: Youqing Wang

Copyright © 2018 Xin Zhang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

A new residual life prediction method for complex systems based on Wiener process and evidential reasoning is proposed to predict the residual life of complex systems effectively. Moreover, the better maintenance strategies and decision supports are provided. For the residual life prediction of complex systems, the maximum likelihood method is adopted to estimate the drift coefficient, and the Bayesian method is adopted to update the parameters of Wiener process. The process of parameters estimation and the probability density function (PDF) of the residual life are deduced. To improve the accuracy of the residual life prediction results, the evidential reasoning (ER) is used to integrate the prediction results of Wiener process. Finally, a case study of gyroscope is examined to illustrate the feasibility and effectiveness of the proposed method, compared with fuzzy theory, which provides an important reference for the optimization of the reliability of complex systems and improvement.

1. Introduction

Nowadays, the residual life prediction is becoming an attention and hot topic in engineering practice. Especially in the performance study of pipeline, oil depot, engine, LED lamp control system, storage tanks, gyroscope, and other complex systems, the residual life prediction has been widely applied. If the residual life of complex systems can be estimated at the beginning of the performance degradation of the systems, and if the performance can be evaluated effectively [1], especially when no major damage has been caused, the happening of unnecessary faults will be avoided to some extent and the maintenance costs will be reduced greatly [2]. Therefore, it is essential to study the residual life prediction of complex systems.

In the traditional life experiment, failure data can be used to predict the life of equipment and device. On the one hand, many failure data cannot be obtained in a short time [3]. On the other hand, some of the systems' characteristics will degrade over time. However, a large amount of information related to reliability and residual life is included in the degradation data. Therefore, the degradation data can be

used to predict the residual life of complex systems. Through the acceleration test, both reliability and life information under different environmental conditions can be obtained. At present, the commonly used acceleration equations include Arrhenius model, Eyring model, Power Law Inverse (IPL), and Hallberg-Peck model. Among them, Arrhenius and Eyring models are always used for thermal accelerated stress in the systems, while temperature is the major factor causing aging. IPL is always used for other different kinds of accelerated stresses except thermal accelerated stress. Hallberg-Peck model synthesizes temperature and humidity that can describe the aging test more accurately under conditions of temperature and humidity.

In current study, the prediction methods can be classified as the physics-based prediction method [4] and data-driven prediction method [5]. With the development of related technologies such as signal acquisition and signal processing, abundant data of system operation can be obtained. According to these data, the corresponding mathematical model is utilized. The method is mainly composed of two kinds of techniques, which are named as artificial intelligence and probability statistics. The artificial intelligence methods have

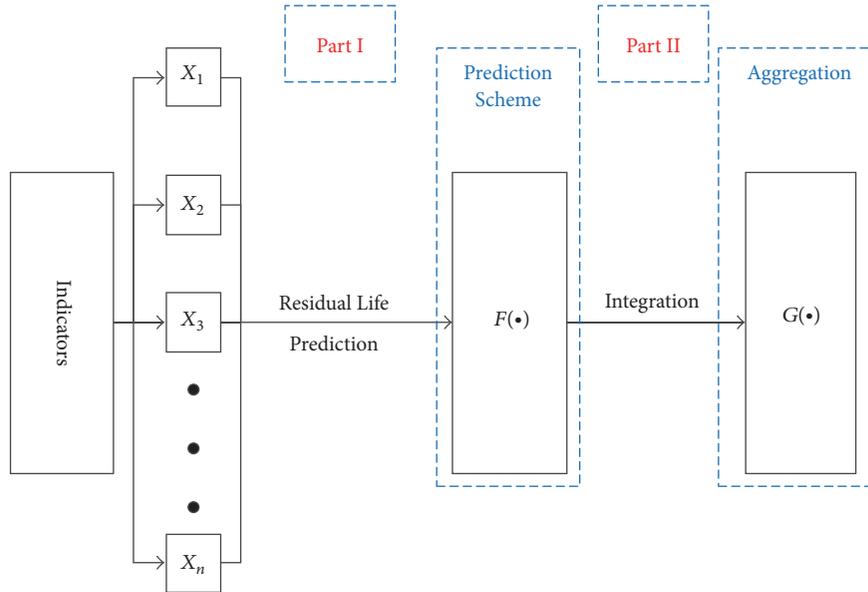


FIGURE 1: Main structure of the new residual life prediction model.

a higher degree of the data fitting, but they cannot obtain the uncertainty of the prediction. The probability statistics can predict the future state better.

Firstly, Wiener process is a kind of the most commonly used probability statistical model, which can describe the degradation process accurately [6, 7]. The degradation model based on Wiener process has obvious advantages in mathematics [8]. Wiener process has been widely applied in many fields [9–11]. Wang proposed a reliability assessment method that fused both prior and on-site degradation data [12]. To update the parameters of Wiener process, the prior data was used to determine the parameters, and the on-site data was integrated by Bayesian method. In [13], the degradation of aeroengine was modeled based on multistage Wiener process. Based on the historical degradation data and the historical failure time data, the prior distribution of the model parameters was estimated by using the expectation maximization (EM) algorithm. In [14], the recursive filtering algorithm and EM algorithm were combined to update the parameters of the Wiener process, and the method was applied to the inertial navigation system to predict the residual life accurately.

Secondly, the evidential reasoning rule (ER rule) proposed by Yang and Xu [15] in 2013 is also a kind of probabilistic statistical model. As a multiattribute decision-making method, the ER rule is widely used to solve the fuzziness and uncertainty of parameters in the evaluation and decision-making process. In [16, 17], the authors have described how to handle the multiattribute decision analysis under uncertainty by the ER rule that is developed based on the decision theory and D-S theory of evidence [15]. Meanwhile, the ER rule is suitable for solving and expressing uncertainty [18–20].

To solve above problems, the prediction framework of the residual life of complex systems is constructed, and the corresponding model of complex systems based on Wiener process and evidential reasoning is established in

this paper. The other contents of this paper are mainly organized as follows. In Section 2, the structure of the residual life prediction model of complex systems is constructed, and the problems to be solved are described in detail. The residual life prediction model of complex systems based on Wiener process and evidential reasoning is established in Section 3, and the details of modeling steps are organized and summarized systematically. In the Section 4, a case study about the residual life prediction of gyroscope is done, based on fuzzy theory and evidential reasoning, respectively, and the simulation results are analyzed. Finally, the practicability of the prediction methods proposed and future work are summarized in the Section 5, as well as the reference of other similar products.

2. New Residual Life Prediction Model and Problem Formulation

2.1. Structure of the New Residual Life Prediction Model. The residual life of complex systems is influenced by many factors. They can be divided into two categories, which are named as the internal factors and the external factors, respectively. The internal factors are composed of materials, structure, and so on, while the external factors include debugging, operation, and maintenance. Thus, it is hard and complex to establish an accurate and complete mathematical prediction model of residual life of complex systems [21]. However, the residual life of the systems can be reflected by some so-called health indicators [22]. Suppose x_i denotes the i th health indicator for the monitoring, $i = 1, 2, \dots, n$. By integrating all the health indicators, a new prediction model of residual life is established, whose structure is shown as Figure 1.

Figure 1 shows the structure of the new model that mainly includes two parts. In the first part, the residual life prediction

model is established based on the selected health indicators and Wiener process. In the second part, the prediction results are integrated by the ER rule to obtain more accurate prediction results.

2.2. Problem Description for Residual Life Prediction. Based on Figure 1, the following three problems should be studied in order to predict the residual life of complex systems.

Problem 1. In Figure 1, the prediction of residual life is necessary. Thus, it is essential to establish a prediction model to get the indicators' prediction value with the obtained indicators' information. Therefore, Problem 1 concentrates on how to establish the following model:

$$\bar{y}_i(t) = F(x_i(1), x_i(2), \dots, x_i(n)), \quad (1)$$

where $i = 1, 2, \dots, n$. $\bar{y}_i(t)$ denotes the prediction value of the residual life of indicator x_i , which is a time series in nature. $F(\bullet)$ represents a nonlinear function.

Problem 2. In practice, in order to represent the relative importance of the health indicators, it is of much necessity to allocate an adaptive weight to each indicator. Thus, Problem 2 concentrates on how to establish the following model to adjust the weight and update the weight coefficient according to the whole available observation data of all indicators:

$$\omega_i(k) = \Psi(x_i(1), x_i(2), \dots, x_i(n)), \quad (2)$$

where $\sum_{i=1}^n \omega_i(k) = 1$. $\Psi(\bullet)$ represents a nonlinear function to determine the weighting coefficient.

Problem 3. To obtain the comprehensive prediction results, it is necessary to integrate various indicators and reprocess the prediction value of the residual life. Therefore, Problem 3 concentrates on how to establish the following integration model to obtain the residual life level of complex systems:

$$g(i) = G(\omega_i(1), \omega_i(2), \dots, \omega_i(n), \bar{y}_i(t)), \quad (3)$$

where $G(\bullet)$ represents a nonlinear function.

Based on the above three problems, the concrete models are shown as follows.

3. Residual Life Prediction Model

In this paper, the following two models are used to solve the above three problems. Firstly, based on Wiener process, a degradation model of complex systems is established to analyze and describe the performance degradation process. In order to eliminate the tendency items, the incremental degradation is adopted. Secondly, the parameters of the degradation model are estimated and updated, with the maximum likelihood method to estimate the parameters, and the Bayesian method to update the parameters. Thirdly, the residual life prediction model of complex systems is established based on evidential reasoning, compared with fuzzy theory, and the prediction results of Wiener process are integrated to obtain more accurate prediction results.

3.1. Prediction Model of Residual Life of Complex Systems Based on Wiener Process

3.1.1. Degradation Model of Complex Systems. Wiener process has been widely used in the modeling of degradation process because of the excellent capability of analyzing and describing the process of degradation [23, 24]. So, it is used to model the performance of complex systems in this paper.

If $Z(t)$ is the degradation measure of the performance of the systems at time t , the degradation model of Wiener process based on the drift parameter is

$$\Delta Z(t) = Z(t + \Delta t) - Z(t) = \theta \Delta t + \varepsilon \Delta W(t), \quad (4)$$

where θ is drift coefficient, ε is diffusion coefficient, Δt is a short time interval, and $W(t)$ is the Standard Brownian Motion which has two features as follows.

Feature 1. $W(t) \sim N(0, t)$.

Feature 2. For any two different times t_1 and t_2 , $t_1 < t_2$, $t_1 + \Delta t < t_2$, $\Delta W(t_1)$, and $\Delta W(t_2)$ are independent.

According to the characteristic of $W(t)$, it is assumed that $\Delta Z(t)$ obeys the normal distribution, that is, $\Delta Z(t) \sim N(\theta \Delta t, \varepsilon^2 \Delta t)$. So, its degradation amount mean is $E(\Delta Z(t)) = \theta \Delta t$, and degradation amount variance is $D(\Delta Z(t)) = \varepsilon^2 \Delta t$. The mean reflects the trend of the degradation data affected by the accelerated stress. So, it is considered that the drift coefficient θ is related to the acceleration stress. The variance reflects the degree of deviation between the degenerate data and the mean, which is usually unrelated to the acceleration stress. So it is considered that the diffusion coefficient ε has nothing to do with the acceleration stress. In this paper, the random parameters θ and ε^2 are used to describe the individual difference of complex systems. Since there is no sufficient prior information of the parameters to be estimated, the prior distribution without information is adopted.

The failure threshold of the performance of some complex systems is often a fixed value, which is set as a constant C . When the current degradation data reaches the threshold value for the first time, the system is failed. So, its residual useful life can be defined as

$$\gamma = \inf \{t \mid \Delta Z(t) \geq C, t > 0\}. \quad (5)$$

It is proved that its first failure threshold time γ is subject to inverse Gaussian distribution, and its probability density function is

$$f(t \mid \theta, \varepsilon) = \frac{C - \Delta Z(t)}{\sqrt{2\pi\varepsilon^2 t^3}} \exp \left[-\frac{(C - \Delta Z(t) - \theta t)^2}{2\varepsilon^2 t} \right]. \quad (6)$$

And the cumulative distribution function is

$$F(t) = \phi \left(\frac{\theta t - \Delta Z(t) - C}{\varepsilon \sqrt{t}} \right) + \exp \left(\frac{2\theta C}{\varepsilon^2} \right) \phi \left[\frac{-(\theta t - \Delta Z(t)) + C}{\varepsilon \sqrt{t}} \right], \quad (7)$$

where $\phi(\bullet)$ is probability density function of the standard normal distribution.

In the accelerated degradation test, temperature and humidity are the most common accelerated stresses that can aggravate the reaction to degenerate the product. Both temperature and humidity are important factors for reliability of some complex systems. High temperature and humid environment may make the service life decrease, so the entire life of the system is seriously affected. In this way, the Hallberg-Peck acceleration model is used to construct the relationship between the drift coefficient with the temperature and humidity stresses. Considering the influence of temperature and humidity comprehensively, the Hallberg-Peck model can describe the accelerated degradation test of the product under the condition of temperature and humidity accurately. Its expression is

$$\theta = A (H_R)^{-3} \exp \left[\frac{E}{K (273 + T)} \right], \quad (8)$$

where A is a constant, $A > 0$, H_R is relative humidity, and E is activation energy in the unit of electron volts (eV). K is Boltzmann constant whose value is 8.617×10^{-5} eV/°C. So, the unit of E/K is °C. In addition, T is relative temperature. The degradation model can be obtained through substituting (8) into (4) as

$$\Delta Z(t, T, H_R) = A (H_R)^{-3} \exp \left[\frac{E}{K \cdot (273 + T)} \right] \Delta t + \varepsilon \Delta W(t) \quad (9)$$

3.1.2. The Estimation of Parameters. The accelerated stresses in the accelerated degradation test are set to T_1, T_2, \dots, T_l . There are six samples for each acceleration stress, and the measured time gradient of each sample is t_1, t_2, \dots, t_n . The degradation amount of the j th sample under the T_i stress is Z_{ijk} at time t_k , and the performance degradation is 0 at the initial time; $\Delta Z_{ijk} = Z_{ijk} - Z_{ij(k-1)}$ is the increment of degradation. By the nature of $W(t)$, $\Delta Z_{ijk}(t) \sim N(\theta_{ij} \Delta t_{ijk}, \varepsilon^2 \Delta t_{ijk})$, we have

$$\begin{aligned} E(\Delta Z_{ijk}) &= \theta_{ij} \Delta t_{ijk} \\ &= A (H_R)^{-3} \exp \left[\frac{E}{K \cdot (273 + T)} \right] \Delta t_{ijk} \end{aligned} \quad (10)$$

$$D(\Delta Z_{ijk}) = \varepsilon^2 \Delta t_{ijk}. \quad (11)$$

In the period of Δt_k , the probability density function (PDF) of the performance degradation data is

$$D(\Delta Z_{ijk}) = \varepsilon^2 \Delta t_{ijk}. \quad (12)$$

The likelihood function of θ and ε^2 can be obtained in

$$L(A, E, \varepsilon) = \prod_{i=1}^l \prod_{j=1}^m \prod_{k=1}^n \frac{1}{\sqrt{2\pi\varepsilon^2\Delta t_{ijk}}} \exp \left[-\frac{(\Delta Z_{ijk} - \theta\Delta t_{ijk})^2}{2\varepsilon^2\Delta t_{ijk}} \right]. \quad (13)$$

The maximum likelihood estimation can be used to estimate the values of θ and ε^2 in (13). But the solving process is very complicated. Therefore, the prior distribution of the parameters is determined firstly, and the parameters are updated by the Bayesian method secondly; then the distribution of the parameters can be obtained. Based on the Bayesian theory, the posterior distribution can be represented as

$$p(\theta | y) \propto f(y | \theta) p(\theta), \quad (14)$$

where $p(\theta | y)$ is the probability density function of posterior distribution, $f(y | \theta)$ is likelihood function, and $p(\theta)$ is the probability density function of prior distribution.

If the distribution of the measured data of the first set sample in stress T_1 is a normal distribution, where the parameters of θ_1 and ε_1^2 are noninformative prior distribution, the prior distribution probability density function can be expressed as

$$p(\theta, \varepsilon^2) \propto \frac{1}{\varepsilon^2}. \quad (15)$$

Let $\theta_a = \theta_1 \Delta t_{11k}$, $\varepsilon_a^2 = \varepsilon_1^2 \Delta t_{11k}$, then $\varepsilon_a^2 = \varepsilon_1^2 \Delta t_{11k}$, the joint posterior distribution of $(\theta_a, \varepsilon_a^2)$ is:

$$\begin{aligned} p(\theta_a, \varepsilon_a^2 | \Delta Z_{11k}) \\ \propto (\varepsilon_a^2)^{-n/2-1} \exp \left[\sum_{k=1}^n -\frac{1}{2\varepsilon_a^2} (\Delta Z_{11k} - \theta_a)^2 \right] \end{aligned} \quad (16)$$

First of all, to update the parameters of the estimated parameters θ_a , ε_a^2 can be regarded as redundant parameters. To integrate the posterior distribution of the integral, the excess parameters are removed, shown in

$$\begin{aligned} p(\theta | \Delta Z_{11k}) \\ \propto \int_0^\infty (\varepsilon_a^2)^{-n/2-1} \exp \left[\sum_{k=1}^n -\frac{1}{2\varepsilon_a^2} (\Delta Z_{11k} - \theta_a)^2 \right] d\varepsilon_a^2. \end{aligned} \quad (17)$$

The equation is finished as

$$\begin{aligned} p(\theta_a | \Delta Z_{11k}) \\ \propto \frac{\Gamma(n/2)}{\left[\sum_{k=1}^n (\Delta Z_{11k} - \theta_a)^2 / 2 \right]^{n/2}} \cdot \int_0^\infty \frac{\left[\sum_{k=1}^n (\Delta Z_{11k} - \theta_a)^2 / 2 \right]^{n/2}}{\Gamma(n/2)} (\varepsilon_a^2)^{-n/2-1} \exp \left[\sum_{k=1}^n -\frac{1}{2\varepsilon_a^2} (\Delta Z_{11k} - \theta_a)^2 \right] d\varepsilon_a^2 \end{aligned}$$

$$= \frac{\Gamma(n/2)}{\left[\sum_{k=1}^n (\Delta Z_{11k} - \theta_a)^2 / 2\right]^{n/2}}. \quad (18)$$

$\sum_{i=1}^n (\Delta Z_{11k} - \theta_a)^2$ is transformed as $n(\Delta Z_{11k} - \theta_a)^2 + \sum_{k=1}^n (\Delta Z_{11k} - \Delta \bar{Z}_{11k})^2$ and then put it into (18):

$$\begin{aligned} p(\theta_a | \Delta Z_{11k}) &\propto \frac{\Gamma(n/2)}{\left[\sum_{k=1}^n (\Delta Z_{11k} - \theta_a)^2 / 2\right]^{n/2}} \\ &\propto \frac{\Gamma(n/2)}{\left[(1/2) \sum_{k=1}^n (\Delta Z_{11k} - \Delta \bar{Z}_{11k})^2\right]^{n/2} \cdot \left[1 + n(\Delta \bar{Z}_{11k} - \theta_a)^2 / \sum_{k=1}^n (\Delta Z_{11k} - \Delta \bar{Z}_{11k})^2\right]^{n/2}} \\ &\propto \left[1 + \frac{n(\Delta \bar{Z}_{11k} - \theta_a)^2}{\sum_{k=1}^n (\Delta Z_{11k} - \Delta \bar{Z}_{11k})^2}\right]^{-n/2} \propto \left[1 + \frac{(\theta_a - \Delta \bar{Z}_{11k})^2}{(n-1)S^2/n}\right]^{-n/2}, \end{aligned} \quad (19)$$

where $\Delta \bar{Z}_{11k} = (1/n) \sum_{k=1}^n \Delta Z_{11k}$, $S^2 = (1/(n-1)) \sum_{k=1}^n (\Delta Z_{11k} - \Delta \bar{Z}_{11k})^2$. From (19), it can be seen that the marginal posterior distribution of θ_1 follows the mean value of $\Delta \bar{Z}_{11k}$,

and the scale parameter fits the normal distribution of S^2/n .

Update the estimated parameter ε_a^2 in

$$\begin{aligned} p(\varepsilon_a^2 | \Delta Z_{11k}) &\propto \int_0^\infty (\varepsilon_a^2)^{-n/2-1} \exp\left[-\frac{1}{2\varepsilon_a^2} \sum_{k=1}^n (\Delta Z_{11k} - \theta_a)^2\right] d\theta_a \\ &\propto (\varepsilon_a^2)^{-n/2-1} \int_0^\infty \exp\left[-\frac{1}{2\varepsilon_a^2} \left(n(\Delta Z_{11k} - \theta_a)^2 + \sum_{k=1}^n (\Delta Z_{11k} - \Delta \bar{Z}_{11k})^2\right)\right] d\theta_a \\ &\propto (\varepsilon_a^2)^{-n/2-1} \int_0^\infty \exp\left[-\frac{(\Delta Z_{11k} - \theta_a)^2}{2\varepsilon_a^2/n} - \frac{S^2}{2\varepsilon_a^2/(n-1)}\right] d\theta_a \\ &\propto (\varepsilon_a^2)^{-n/2-1} \cdot \exp\left[-\frac{(n-1)S^2}{2\varepsilon_a^2}\right] \cdot \int_{-\infty}^{+\infty} \exp\left[-\frac{(\Delta \bar{Z}_{11k} - \theta_a)^2}{2\varepsilon_a^2/n}\right] d\theta_a \propto (\varepsilon_a^2)^{-n/2-1} \cdot \exp\left[-\frac{(n-1)S^2}{2\varepsilon_a^2}\right]. \end{aligned} \quad (20)$$

From (20), it can be seen that the posterior distribution of ε_a^2 is proportional to the probability density function of the inverse Gamma distribution. The shape parameter is $(n-1)/2$ and the scale parameter is $(n-1)S^2/2$. Based on (19) and (20), the mean value of the posterior distribution of θ_1 is $\Delta \bar{Z}_{11k}/\Delta t_{11k}$, and the scale parameter is $s^2 \Delta t_{11k}/n$. The shape parameter of the posterior distribution of ε_1^2 is $(n-1)/2$, and the scale parameter is $(n-1)S^2/2\Delta t_{11k}$. The estimated values of the parameters $\hat{\theta}_1$ and $\hat{\varepsilon}_1^2$ under the stress of T_1 are

$$\begin{aligned} \hat{\theta}_1 &= \frac{1}{m} \sum_{j=1}^m \theta_{1j}; \\ \hat{\varepsilon}_1^2 &= \frac{1}{m} \sum_{j=1}^m \varepsilon_{1j}^2. \end{aligned} \quad (21)$$

In the same way, the estimated parameters can be obtained under T_2, T_3, \dots, T_i .

To T_i , evaluate the logarithmic of (10):

$$\ln \theta_i = \ln A - 3 \ln(H_{Ri}) + \frac{E}{K(273 + T_i)}. \quad (22)$$

So, T_i and T_j ($i = 1, 2, \dots, l; j = 1, 2, \dots, l; i \neq j$) are combined:

$$\begin{aligned} \ln \theta_i - \ln \theta_j &= 3 [\ln(H_{Rj}) - \ln(H_{Ri})] \\ &+ \left[\frac{1}{K(273 + T_i)} - \frac{1}{K(273 + T_j)} \right] E. \end{aligned} \quad (23)$$

Substitute the estimated value of (21) into (23):

$$\hat{E}_{ij} = \frac{(\ln \hat{\theta}_i - \ln \hat{\theta}_j) - 3 [\ln(H_{Rj}) - \ln(H_{Ri})]}{1/K(273 + T_i) - 1/K(273 + T_j)}. \quad (24)$$

The estimated value of E can be calculated as

$$\hat{E} = \frac{1}{l(l-1)} \sum_{i=1}^l \sum_{j=1, i \neq j}^l \hat{E}_{ij}. \quad (25)$$

Based on (8), the estimated value of A is

$$\hat{A} = \frac{1}{l} \sum_{i=1}^l \frac{\hat{\theta}_i}{(H_{Ri})^{-3} \exp[\hat{E}/K(273 + T_i)]}. \quad (26)$$

Since the diffusion coefficient ε is independent of the acceleration stress, its estimated value is

$$\hat{\varepsilon}^2 = \frac{1}{l} \sum_{i=1}^l \hat{\varepsilon}_i^2. \quad (27)$$

After the values of the parameters in the degradation model have been obtained, the mean and variance of the degradation data under normal stress level can be obtained according to the accelerated degradation model. And the corresponding failure distribution can be obtained by substituting them into (6).

3.2. Residual Life Prediction Model Based on Evidential Reasoning. As mentioned in Section 1, considering the advantages of evidential reasoning in theoretical research and practical application, the prediction model of residual life of complex systems based on evidential reasoning is established.

The core of the ER rule is a kind of distributed type with a belief structure. In this way, any input information, whether it is accurate data or inaccurate data with uncertainty, subjective or objective judgment, can be consistently modeled based on the power set. Moreover, the ER rule also handles with the probabilistic uncertainty by integrating the reliability of all indicators in (24). The details are as follows.

(1) Transformation of All Kinds of Input Indicators. On the one hand, different indicators can be described in different ways. On the other hand, there is always ignorance and vagueness in the prediction of residual life of complex systems. In addition, both qualitative and quantitative information are considered

with uncertainties [25]. So, it is of much importance to transform various types of health indicators into belief distribution.

Define N distinctive residual life grades as represented by

$$H = \{H_1, H_2, \dots, H_i, \dots, H_N\}, \quad (28)$$

where H_i is the i th residual life grade. It is assumed that H_{i+1} is inferior to H_i . In addition, H could be determined by circumstances, including special conditions and personal experience. For example, three grades "Long" (H_1), "Average" (H_2), and "Short" (H_3) can be chosen in the prediction of residual life of complex systems.

Thus, the distribution can be described as

$$S(x_j) = \{(H_i, \beta_{i,j}), i = 1, 2, \dots, N\}, \quad (29)$$

$$j = 1, 2, \dots, m,$$

where $0 \leq \beta_{i,j}, \sum_{i=1}^N \beta_{i,j} \leq 1$, and $\beta_{i,j}$ denotes a degree of belief. The above distributed assessment reads that the indicator x_j is assessed to the grade H_i with the degree of belief $\beta_{i,j}$, which can be calculated as

$$\beta_{i,j}(x_j) = \frac{U(H_{i+1}) - x_i}{U(H_{i+1}) - U(H_i)},$$

$$U(H_i) \leq x_i \leq U(H_{i+1}) \quad (30)$$

$$\beta_{i+1,j}(x_j) = 1 - \beta_{i,j}(x_j), \quad U(H_i) \leq x_j \leq U(H_{i+1})$$

$$\beta_{k,j}(x_j) = 0, \quad k = 1, 2, \dots, N, \quad k \neq i, i+1,$$

where $k = 1, 2, \dots, N-1$, $U(H_i)$ denotes a quantitative value that can be judged as a referential value H_i .

(2) Integration of Multi-Indicators. For complex systems, the residual life at time t relies on all health indicators, which means the residual life can be obtained through multiple health indicators. The information in (28) can be integrated by the ER rule. The results of the residual life prediction $z(t)$ can be expressed as follows:

$$L(z(t)) = \{(H_i, \beta_i(t)), i = 1, 2, \dots, N\}, \quad (31)$$

where the distribution $L(z(t))$ represents the assessment of the residual life of complex systems at time t , from which the detailed life grades and belief degrees can be seen clearly. $\beta_i(t)$, $i = 1, 2, \dots, N$ can be calculated as follows [15]:

$$\beta_i(t) = \frac{\mu(t) \times [\prod_{l=1}^n (\bar{\omega}_l(t) \beta_{i,l}(t) + 1 - \bar{\omega}_l(t) \sum_{k=1}^N \beta_{k,l}(t))] - \mu(t) \times [\prod_{l=1}^n (1 - \bar{\omega}_l(t) \sum_{k=1}^N \beta_{k,l}(t))]}{1 - \mu(t) \times [\prod_{l=1}^n (1 - \bar{\omega}_l(t))]} - \frac{\mu(t) \times [\prod_{l=1}^n (1 - \bar{\omega}_l(t) \sum_{k=1}^N \beta_{k,l}(t))]}{1 - \mu(t) \times [\prod_{l=1}^n (1 - \bar{\omega}_l(t))]}$$

$$\mu(t) = \left[\sum_{i=1}^N \prod_{l=1}^n \left(\bar{\omega}_l(t) \beta_{i,l}(t) + 1 - \bar{\omega}_l(t) \sum_{k=1}^N \beta_{k,l}(t) \right) - (N-1) \prod_{l=1}^n \left(1 - \bar{\omega}_l(t) \sum_{k=1}^N \beta_{k,l}(t) \right) \right]^{-1} \quad (32)$$

$$\bar{\omega}_l(t) = \frac{\omega_l(t)}{1 + \omega_l(t) - r_l},$$

where $\omega_l(t)$ denotes the weight coefficient of the l th indicator at time t and r_l denotes the reliability of the l th indicator.

(3) *Utility Evaluation.* In order to visually represent the distributed assessment results, the expected utility is introduced. $u(H_i)$, $i = 1, 2, \dots, N$ is the utility of evaluation grade H_i , and the expected utility of the prediction result $L(z(t))$ is defined as follows:

$$u(L(z(t))) = \sum_{i=1}^N \beta_i(t) u(H_i). \quad (33)$$

(4) *Ways to Calculate the Adaptive Weighting Coefficient.* The weighting coefficient represents the importance of the indicators for prediction results [15, 26]. In this paper, the adaptive weight is calculated with CVBW method [27] shown as follows.

Suppose the observed data of indicator x_j within t is denoted by $z_j(k)$, $k = 1, 2, \dots, t$, whose mean value is \bar{z}_j . So, \bar{z}_j can be calculated by

$$\bar{z}_j = \frac{1}{t} \sum_{k=1}^t z_j(k). \quad (34)$$

s_{z_j} denotes the mean square error of the observed data of indicator x_j within time t , which can be calculated by

$$s_{z_j} = \left[\frac{1}{t-1} \sum_{k=1}^t (z_j(k) - \bar{z}_j)^2 \right]^{1/2}. \quad (35)$$

Also, there is

$$v_{z_j} = \frac{s_{z_j}}{\bar{z}_j}. \quad (36)$$

The weighting coefficient of x_j within t can be calculated by

$$\omega_j(t) = \frac{v_{z_j}}{\sum_{j=1}^n v_{z_j}}. \quad (37)$$

3.3. A New Algorithm for Residual Life Prediction of Complex Systems Based on Wiener Process and Evidential Reasoning. Based on the above analysis, a new algorithm for the residual life prediction of complex systems based on Wiener process and evidential reasoning is proposed in this section. The details are shown as follows.

Step 1. Construct the framework of the residual life prediction model of complex systems, and comb through the problems to be solved.

Step 2. Based on the working mechanism of complex systems, establish the degradation model of the systems, as is shown in (9).

Step 3. Estimate and update the parameters of the degradation model, using the maximum likelihood method and Bayesian method, respectively, as is shown in (21) and (27).

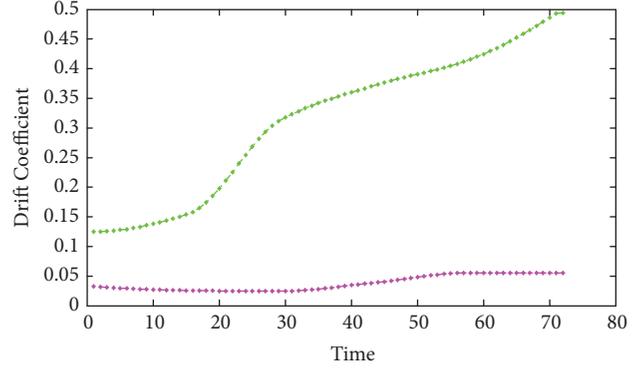


FIGURE 2: Degradation of the drift coefficient of gyroscope.

Step 4. Select the life indicators of complex systems, and obtain the residual life of complex systems based on Wiener process.

Step 5. Establish the residual life prediction model of complex systems based on evidential reasoning, as is shown in the Section 3.2.

Step 6. Calculate the adaptive weight, as is shown in (37). Based on the ER rule, make a precise prediction for the residual life prediction results of Wiener process.

4. Case Study

To prove the practicability of the proposed prediction models of complex systems, the residual life of gyroscope is studied in this paper. Considering the fact that the residual life of gyroscope is mainly determined by its drift coefficient which changes as time, the drift coefficient for zero and once term are chosen to measure the residual life of gyroscope.

4.1. Residual Life Prediction of Gyroscope Based on Wiener Process. Under the condition of constant temperature 20°C and humidity 70%, the drift coefficient of the gyroscope is tested once a month, and the degradation values are obtained. To get more accurate results, the test is repeated. Figure 2 shows the results.

In Figure 2, K_{0x} denotes the drift coefficient for zero item and K_{1x} denotes the drift coefficient for once item. Obviously, the two groups of degradation data are roughly linear, random processes, whose increments are not strict. Thus, the degradation process can be determined in Wiener process. In order to fit the data, the increment is introduced based on (4). According to the characteristics of Wiener process, $\Delta Z(t) \sim N(\theta\Delta t, \varepsilon^2\Delta t)$, it can be described by the probability distribution $\Delta Z(t)$. The fitting results are shown as Figures 3 and 4.

As can be seen from Figures 3 and 4, the drift coefficient degradation process of gyroscope is a Wiener process. Using the method described in the Section 3.1.2, four groups of degenerate data are updated to obtain the normal inverse Gamma posterior distribution parameters of θ and ε^2 , which are shown in Table 1.

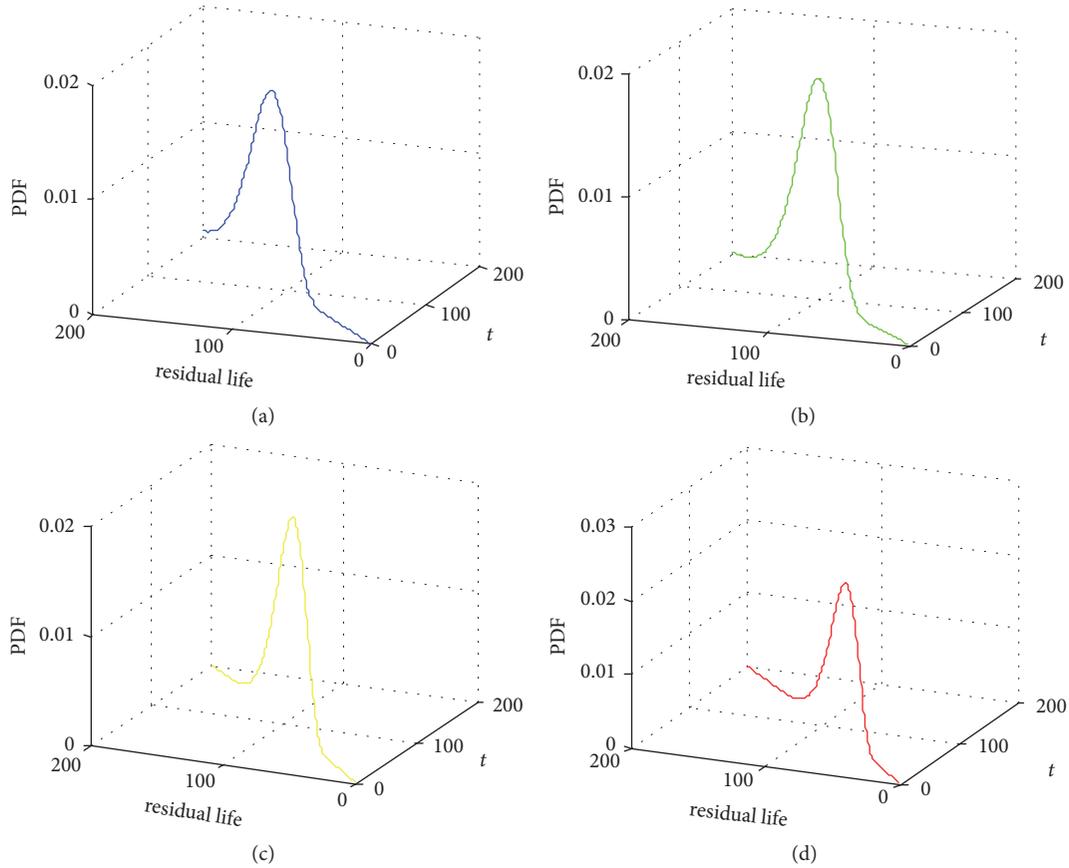


FIGURE 3: The degradation of K_{ox} of gyroscope.

TABLE 1: The results of the estimated parameters.

| Estimated parameters | Drift coefficient $\hat{\theta}$ | Diffusion coefficient ϵ^2 |
|----------------------|----------------------------------|------------------------------------|
| K_{ox} | 0.0032 | 0.0072 |
| K_{sx} | 0.0026 | 0.0018 |

Remark 4. The parameters are set as shown in Table 2.

Finally, the prediction results of the residual life of gyroscope based on K_{ox} and K_{sx} are obtained, which are shown in Tables 3 and 4, respectively.

According to Tables 3 and 4, the prediction results of the residual life of gyroscope based on K_{ox} and K_{sx} are close to the real value, which has proved the accuracy of Wiener process in this paper. Therefore, the method proposed in the Section 3.1 can provide a reference for predicting the residual life of gyroscope in engineering practice.

Remark 5. Based on the actual degradation of gyroscope, the residual life of gyroscope is generally thought to be linear decreasing tendency.

4.2. Residual Life Prediction of Gyroscope Based on Evidential Reasoning. Based on evidential reasoning, the prediction results of the residual life of Wiener process are integrated. The details are as follows.

Firstly, calculate the weighting coefficient. The results are shown in Table 5.

Secondly, calculate the evidence reliability. Based on expert knowledge, the reliability of K_{ox} and K_{sx} are 0.8262 and 1, respectively. In addition, the assessment grades are set as “Very Long” (H_1), “Long” (H_2), “Average” (H_3), “Short” (H_4), and “Very Short” (H_5), shown as follows.

$$\begin{aligned}
 H &= \{H_1, H_2, H_3, H_4, H_5\} \\
 &= \{\text{Very Long, Long, Average, Short, Very Short}\}.
 \end{aligned}
 \tag{38}$$

In general, if the total life of gyroscope is 200 hours, then H_1 denotes the residual life for 200 hours, H_2 for 150, H_3 for 100, H_4 for 50, and H_5 for 0.

The standard values of each grade are 200, 150, 100, 50, and 0, respectively, denoted by $A, B, C, D,$ and E . Hence, the evidence is shown in Table 6.

Thirdly, integrate the evidence. Based on the method in the Section 3.2, the integration results are shown in Table 7.

It can be seen from Table 7 that when the initial value of the degradation of the gyroscope is 0.10, the probability that the residual life of gyroscope is in the state H_1 is 0.0283, and the probability of being in state H_2 is 0.9717. So, the belief distribution can be described as follows:

$$\begin{aligned}
 L(z(0.10)) &= \{(H_1, 0.0283), (H_2, 0.9717), (H_3, 0), \\
 &\quad (H_4, 0), (H_5, 0)\}.
 \end{aligned}
 \tag{39}$$

TABLE 2: Parameters setting.

| Parameters | Initial value of degradation | Threshold value | Maximum life |
|------------|------------------------------|-----------------|--------------|
| (a) | 0.10 | | |
| (b) | 0.15 | 0.5 | 200 |
| (c) | 0.20 | | |
| (d) | 0.25 | | |

TABLE 3: Prediction values of residual life based on K_{ox} compared with the real value under different initial values of degradation.

| Initial value of degradation | Real value | Predictive value | PDF |
|------------------------------|------------|------------------|--------|
| 0.10 | 160 | 120 | 0.0166 |
| 0.15 | 140 | 104 | 0.0179 |
| 0.20 | 120 | 88 | 0.0195 |
| 0.25 | 100 | 72 | 0.0216 |

TABLE 4: Prediction values of residual life based on K_{sx} compared with the real value under different initial values of degradation.

| Initial value of degradation | Real value | Predictive value | PDF |
|------------------------------|------------|------------------|--------|
| 0.10 | 160 | 153 | 0.0462 |
| 0.15 | 140 | 134 | 0.0494 |
| 0.20 | 120 | 115 | 0.0533 |
| 0.25 | 100 | 95 | 0.0585 |

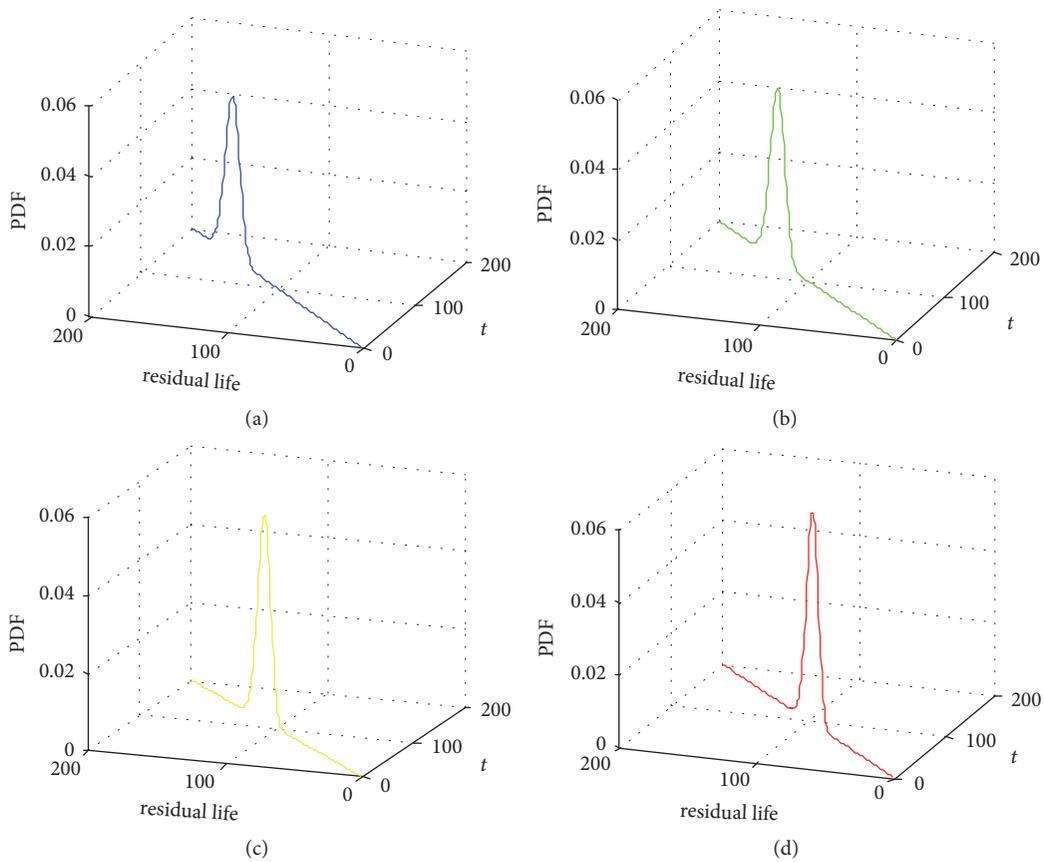


FIGURE 4: The degradation of K_{sx} of gyroscope.

TABLE 5: The weighting coefficient of residual life of *Kox* and *Ksx*.

| Indicators (x_j) | Mean value (\bar{z}_j) | Mean square error (s_{z_j}) | Proportion (v_{z_j}) | Weighting coefficient ($\omega_j(t)$) |
|----------------------|----------------------------|---------------------------------|--------------------------|---|
| <i>Kox</i> | 96 | 20.6559 | 0.2152 | 0.5177 |
| <i>Ksx</i> | 124.25 | 24.9182 | 0.2005 | 0.4823 |

TABLE 6: Two pieces of independent evidence.

| Initial value of degradation | $p_{\theta,i}$ | A | B | C | D | E |
|------------------------------|----------------|------|------|------|------|---|
| 0.10 | <i>Kox</i> | 0 | 0.4 | 0.6 | 0 | 0 |
| | <i>Ksx</i> | 0.06 | 0.94 | 0 | 0 | 0 |
| 0.15 | <i>Kox</i> | 0 | 0.08 | 0.92 | 0 | 0 |
| | <i>Ksx</i> | 0 | 0.68 | 0.32 | 0 | 0 |
| 0.20 | <i>Kox</i> | 0 | 0 | 0.76 | 0.24 | 0 |
| | <i>Ksx</i> | 0 | 0.3 | 0.7 | 0 | 0 |
| 0.25 | <i>Kox</i> | 0 | 0 | 0.44 | 0.56 | 0 |
| | <i>Ksx</i> | 0 | 0 | 0.1 | 0.9 | 0 |

TABLE 7: The integration results.

| Initial value of degradation | Prediction results | H_1 | H_2 | H_3 | H_4 | H_5 |
|------------------------------|--------------------|--------|--------|--------|--------|-------|
| 0.10 | 120 | 0.0229 | 0.9771 | 0 | 0 | 0 |
| | 153 | | | | | |
| 0.15 | 104 | 0 | 0.3654 | 0.6346 | 0 | 0 |
| | 134 | | | | | |
| 0.20 | 88 | 0 | 0.0911 | 0.9089 | 0 | 0 |
| | 115 | | | | | |
| 0.25 | 72 | 0 | 0 | 0.0862 | 0.9138 | 0 |
| | 95 | | | | | |

Therefore, we have a probability of 0.0283 that the residual life of gyroscope is very long, 0.9717 that the residual life is long, and 0 for other grades.

Similarly, the other results are shown as follows.

$$\begin{aligned}
 L(z(0.15)) &= \{(H_1, 0), (H_2, 0.4130), (H_3, 0.5870), \\
 &\quad (H_4, 0), (H_5, 0)\} \\
 L(z(0.20)) &= \{(H_1, 0), (H_2, 0.1161), (H_3, 0.8839), \\
 &\quad (H_4, 0), (H_5, 0)\} \\
 L(z(0.25)) &= \{(H_1, 0), (H_2, 0), (H_3, 0.0878), \\
 &\quad (H_4, 0.9122), (H_5, 0)\}.
 \end{aligned}
 \tag{40}$$

Therefore, the ER rule has well explained the prediction results of Wiener process, which is consistent with the actual observation results. In the engineering practice, when the drift coefficient of gyroscope is reduced to 0.25, it can be considered that the residual life of gyroscope is very short, and a new gyroscope should be replaced for use.

Finally, all the above methods to predict the residual life of gyroscope are shown in Table 8.

From Table 8, the common residual life prediction results of gyroscope can be obtained by Wiener process, while a specific prediction result can be obtained by the ER rule, which can integrate all of the former results.

To clearly demonstrate the superiority of the proposed methods, traditional fuzzy theory has been used to deal with the residual life prediction results of Wiener process [28–30]. Assuming that the set $U = \{A, B, C, D, E\}$, and that the fuzzy subsets are *Kox* and *Ksx*, respectively, based on Table 6, the membership of each subset is shown in Table 9.

Compared with Table 8, Table 9 shows that the final results obtained by fuzzy theory cannot be explained clearly, which are only some uncertain values. Therefore, the prediction results of the ER rule are proved to be more feasible and effective.

5. Conclusion

In this paper, a new model based on Wiener process and evidential reasoning is proposed to predict the residual life

TABLE 8: The final results by ER.

| Methods | Initial value of degradation | Indicators | Prediction results | | PDF |
|----------------|------------------------------|------------|--------------------|--|--------|
| Wiener process | 0.10 | <i>Kox</i> | 120 | | 0.0166 |
| | | <i>Ksx</i> | 153 | | 0.0462 |
| | 0.15 | <i>Kox</i> | 104 | | 0.0179 |
| | | <i>Ksx</i> | 134 | | 0.0494 |
| | 0.20 | <i>Kox</i> | 88 | | 0.0195 |
| | | <i>Ksx</i> | 115 | | 0.0533 |
| | 0.25 | <i>Kox</i> | 72 | | 0.0216 |
| | | <i>Ksx</i> | 95 | | 0.0585 |

| Methods | Initial value of degradation | Indicators | Prediction results | | | | |
|---------|------------------------------|------------|--------------------|--------|--------|--------|-------|
| | | | H_1 | H_2 | H_3 | H_4 | H_5 |
| ER | 0.10 | <i>Kox</i> | 0.0283 | 0.9717 | 0 | 0 | 0 |
| | | <i>Ksx</i> | | | | | |
| | 0.15 | <i>Kox</i> | 0 | 0.4130 | 0.5870 | 0 | 0 |
| | | <i>Ksx</i> | | | | | |
| | 0.20 | <i>Kox</i> | 0 | 0.1161 | 0.8839 | 0 | 0 |
| | | <i>Ksx</i> | | | | | |
| | 0.25 | <i>Kox</i> | 0 | 0 | 0.0878 | 0.9122 | 0 |
| | | <i>Ksx</i> | | | | | |

TABLE 9: The membership of *Kox* and *Ksx* by fuzzy theory.

| Initial value of degradation | Prediction results | Membership | | | | |
|------------------------------|--------------------|--|---|---|---|---|
| | | A | B | C | D | E |
| 0.10 | 120 | $\left\{ \frac{0}{A} + \frac{0.4}{B} + \frac{0.6}{C} + \frac{0}{D} + \frac{0}{E} \right\}$ | | | | |
| | 153 | $\left\{ \frac{0.06}{A} + \frac{0.94}{B} + \frac{0}{C} + \frac{0}{D} + \frac{0}{E} \right\}$ | | | | |
| 0.15 | 104 | $\left\{ \frac{0}{A} + \frac{0.08}{B} + \frac{0.92}{C} + \frac{0}{D} + \frac{0}{E} \right\}$ | | | | |
| | 134 | $\left\{ \frac{0}{A} + \frac{0.68}{B} + \frac{0.32}{C} + \frac{0}{D} + \frac{0}{E} \right\}$ | | | | |
| 0.20 | 88 | $\left\{ \frac{0}{A} + \frac{0}{B} + \frac{0.76}{C} + \frac{0.24}{D} + \frac{0}{E} \right\}$ | | | | |
| | 115 | $\left\{ \frac{0}{A} + \frac{0.3}{B} + \frac{0.7}{C} + \frac{0}{D} + \frac{0}{E} \right\}$ | | | | |
| 0.25 | 72 | $\left\{ \frac{0}{A} + \frac{0}{B} + \frac{0.44}{C} + \frac{0.56}{D} + \frac{0}{E} \right\}$ | | | | |
| | 95 | $\left\{ \frac{0}{A} + \frac{0}{B} + \frac{0.1}{C} + \frac{0.9}{D} + \frac{0}{E} \right\}$ | | | | |

of complex systems. The relative importance of each health indicator is considered, which provides an effective method for predicting the residual life of complex systems. The model proposed in this paper is mainly composed of two parts that are prediction and integration. Through the case study of gyroscope and a comparison with fuzzy theory, it is proved more feasible and effective in engineering practice. As offline prediction proposed in this paper is unable to precisely and timely reflect the systems' life level, its effectiveness, and validity to deal with online problems and more complex problems should be further tested. At the same time, the reliability

of all health indicators is crucial for the residual life prediction of the whole system and it is influenced by various factors, including the noise, sensor's characteristics, and malfunction. In addition, the correlation between each indicator needs to be considered. Hence, further study on how to obtain the reliability and correlation of indicators by integrating all factors based on online prediction needs to be done in the future.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

- [1] Y. Q. Wang, H. Zhang, S. Wei, D. Zhou, and B. Huang, "Control performance assessment for ILC-controlled batch processes in two-dimensional system framework," *IEEE Transactions on Systems, Man and Cybernetics: Systems*, Article ID 2672563, 2017.
- [2] Y. Wang, D. Zhao, Y. Li, and S. Ding, "Unbiased minimum variance fault and state estimation for linear discrete time-varying two-dimensional systems," *Institute of Electrical and Electronics Engineers Transactions on Automatic Control*, vol. 62, no. 10, pp. 5463–5469, 2017.
- [3] C. S. Guo, Y. F. Zhang, N. Wan, R. Li, H. Zhu, and S. W. Feng, "The investigation of LED degradation model based on the chemical kinetics," *Acta Physica Sinica*, vol. 21, pp. 468–473, 2013.
- [4] L. Liao and F. Köttig, "A hybrid framework combining data-driven and model-based methods for system remaining useful life prediction," *Applied Soft Computing*, vol. 44, pp. 191–199, 2016.
- [5] Z.-Q. Wang, C.-H. Hu, W.-B. Wang, and G.-J. Dong, "Wiener process-based online prediction method of remaining useful life for draught fans in steel mills," *Journal of University of Science and Technology Beijing*, vol. 36, no. 10, pp. 1361–1368, 2014.
- [6] Z.-Q. Wang, C.-H. Hu, W. Wang, and X.-S. Si, "An additive wiener process-based prognostic model for hybrid deteriorating systems," *IEEE Transactions on Reliability*, vol. 63, no. 1, pp. 208–222, 2014.
- [7] Z. Q. Wang, C. H. Hu, W. Wang, Z. J. Zhou, and X. S. Si, "A case, study of remaining storage life prediction using stochastic filtering with the influence of condition monitoring," *Reliability Engineering & System Safety*, vol. 132, no. 12, pp. 186–195, 2014.
- [8] X. Wang, S. Lin, S. Wang, Z. He, and C. Zhang, "Remaining useful life prediction based on the Wiener process for an aviation axial piston pump," *Chinese Journal of Aeronautics*, vol. 29, no. 3, pp. 779–788, 2016.
- [9] K. Hoshino, Y. Nishimura, Y. Yamashita, and D. Tsubakino, "Global asymptotic stabilization of nonlinear deterministic systems using Wiener processes," *Institute of Electrical and Electronics Engineers Transactions on Automatic Control*, vol. 61, no. 8, pp. 2318–2323, 2016.
- [10] Q. Guan, Y. Tang, and A. Xu, "Objective Bayesian analysis accelerated degradation test based on Wiener process models," *Applied Mathematical Modelling: Simulation and Computation for Engineering and Environmental Systems*, vol. 40, no. 4, pp. 2743–2755, 2016.
- [11] Z. Wang, J. Li, Y. Zhang, H. Fu, C. Liu, and S. Krishnaswamy, "A novel Wiener process model with measurement errors for degradation analysis," *Eksploracja i Niezawodność*, vol. 18, no. 3, pp. 396–405, 2016.
- [12] X. Wang, "Wiener processes with random effects for degradation data," *Journal of Multivariate Analysis*, vol. 101, no. 2, pp. 340–351, 2010.
- [13] N. Cucu Laurenciu and S. D. Cotofana, "A nonlinear degradation path dependent end-of-life estimation framework from noisy observations," *Microelectronics Reliability*, vol. 53, no. 9–11, pp. 1213–1217, 2013.
- [14] L. Zhu, H.-F. Zuo, and J. Cai, "Performance reliability prediction for civil aviation aircraft engine based on wiener process," *Journal of Aerospace Power*, vol. 28, no. 5, pp. 1006–1012, 2013.
- [15] J.-B. Yang and D.-L. Xu, "Evidential reasoning rule for evidence combination," *Artificial Intelligence*, vol. 205, pp. 1–29, 2013.
- [16] Z. J. Zhou, L. L. Chang, C. H. Hu, X. X. Han, and Z. G. Zhou, "A new BRB-ER-based model for assessing the lives of products using both failure data and expert knowledge," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 46, no. 11, pp. 1529–1543, 2016.
- [17] Z. J. Zhou, C. H. Hu, G. Y. Hu, X. X. Han, B. C. Zhang, and Y. W. Chen, "Hidden behavior prediction of complex systems under testing influence based on semiquantitative information and belief rule base," *IEEE Transactions on Fuzzy Systems*, vol. 23, no. 6, pp. 2371–2386, 2015.
- [18] L. Chang, Z. Zhou, Y. You, L. Yang, and Z. Zhou, "Belief rule based expert system for classification problems with new rule activation and weight calculation procedures," *Information Sciences*, vol. 336, no. 1, pp. 75–91, 2016.
- [19] Z. J. Zhou, C. H. Hu, J. B. Yang, D. L. Xu, and D. H. Zhou, "Online updating belief-rule-base using the RIMER approach," *IEEE Transactions on Systems, Man and Cybernetics, Part A: Systems and Humans*, vol. 41, no. 6, pp. 1225–1243, 2011.
- [20] Z. J. Zhou, C. H. Hu, and B. C. Zhang, "Hidden behavior prediction of complex systems based on hybrid information," *IEEE Transactions on Cybernetics*, vol. 43, no. 2, pp. 402–411, 2013.
- [21] X. S. Si, C. H. Hu, and Z. J. Zhou, "Fault prediction model based on evidential reasoning approach," *Science China Information Sciences*, vol. 53, no. 10, pp. 2032–2046, 2010.
- [22] V. Gitelman, E. Doveh, and S. Hakkert, "Designing a composite indicator for road safety," *Safety Science*, vol. 48, no. 9, pp. 1212–1224, 2010.
- [23] J. Q. Liu, J. W. Xie, H. F. Zuo, and M. L. Zhang, "Remaining lifetime prediction for aeroengines based on Wiener process with random effects," *Acta Aeronautica et Astronautica Sinica*, vol. 02, pp. 564–574, 2015.
- [24] X.-S. Si, W. Wang, C.-H. Hu, M.-Y. Chen, and D.-H. Zhou, "A Wiener-process-based degradation model with a recursive filter algorithm for remaining useful life estimation," *Mechanical Systems and Signal Processing*, vol. 35, no. 1–2, pp. 219–237, 2013.
- [25] Z. G. Zhou, F. Liu, L. L. Li, Z. J. Zhou, J. B. Yang, and Z. L. Wang, "A cooperative belief rule based decision support system for lymph node metastasis diagnosis in gastric cancer," *Knowledge-Based Systems*, vol. 85, no. 9, pp. 62–70, 2015.
- [26] D.-C. Ahn, J.-W. Lee, S.-J. Shin, and W.-J. Song, "A new robust variable weighting coefficients diffusion LMS algorithm," *Signal Processing*, vol. 131, pp. 300–306, 2017.
- [27] F. J. Zhao, Z. J. Zhou, C. H. Hu, L. L. Chang, Z. G. Zhou, and G. L. Li, "A new evidential reasoning-based method for online safety assessment of complex systems," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, pp. 1–13, 2016.
- [28] J. Q. Zhang, "Sampling for building energy consumption with fuzzy theory," *Energy and Buildings*, vol. 156, pp. 78–84, 2017.
- [29] B. Li, G. Yang, R. Wan, L. Zhang, Y. Zhang, and X. Dai, "Using fuzzy theory and variable weights for water quality evaluation in Poyang Lake, China," *Chinese Geographical Science*, vol. 27, no. 1, pp. 39–51, 2017.
- [30] H.-C. Chang, S.-C. Lin, C.-C. Kuo, and C.-Y. Lin, "Fuzzy Theory-Based Partial Discharge Technique for Operating State Diagnosis of High-Voltage Motor," *International Journal of Fuzzy Systems*, vol. 18, no. 6, pp. 1092–1103, 2016.

Research Article

A New Synergistic Forecasting Method for Short-Term Traffic Flow with Event-Triggered Strong Fluctuation

Darong Huang , Zhenping Deng , and Bo Mi

Institute of Information Science and Engineering, Chongqing Jiaotong University, Chongqing 400074, China

Correspondence should be addressed to Darong Huang; drhuang@cqjtu.edu.cn

Received 22 December 2017; Accepted 12 February 2018; Published 19 March 2018

Academic Editor: Zhijie Zhou

Copyright © 2018 Darong Huang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Directing against the shortcoming of low accuracy in short-term traffic flow prediction caused by strong traffic flow fluctuation, a novel method for short-term traffic forecasting based on the combination of improved grey Verhulst prediction algorithm and first-order difference exponential smoothing is proposed. Firstly, we constructed an improved grey Verhulst prediction model by introducing the Markov chain to its traditional version. Then, based on an introduced dynamic weighting factor, the improved grey Verhulst prediction method, and the first-order difference exponential smoothing technique, the new method for short-term traffic forecasting is completed in an efficient way. Finally, experiment and analysis are carried out in the light of actual data gathered from strong fluctuation environment to verify the effectiveness and rationality of our proposed scheme.

1. Introduction

In recent years, the popularization of seamless links among heterogeneous traffic equipment brought about higher requirements on the real-time and reliability of short-term traffic flow prediction. With continuous improvement of traffic information processing, how to predict the short-term traffic flow accurately and effectively has aroused wide attention of scholars domestically and abroad [1–3], whereupon numerous of prominent research results have emerged. So far, relevant academic circles mainly focus on the construction and optimization of prediction models in terms of time series, linear regression, historical average model, Kalman filtering, grey theory, chaos theory, nonparametric regression, neural network, support vector machine, dynamic traffic assignment model, and so forth [4–9]. These algorithms and models mentioned above are relatively mature, and their prediction effects are acceptable under the environment of favorable traffic flow stability. However, once the traffic data are seriously fluctuated, one single model can neither guarantee the prediction accuracy nor break through certain limitations of operations to predict short-term traffic flow under the environments of heterogeneous traffic equipment transmission.

In order to solve the aforementioned problems, domestic and foreign scholars have contrived an improved model to

realize fusion prediction with the advantages of different models integrated of short-term traffic flow. For example, Xie et al. [10] improved the search efficiency of K -nearest neighbor algorithm by cooperating the multivariate statistical regression model and the pattern distance search method based on the analysis of original K -nearest neighbor algorithm for short-term traffic flow prediction. Their experimental results indicated that the prediction effect is better when the K values are reasonable. Fan et al. [4] exploited the characteristics of time-varying and nonlinearity of traffic flow and proposed a new hybrid forecasting model based on the nonparametric regression model and the BP neural network model; they also employed fuzzy control to determine the weight of each single model. However, the basic support platform of the hybrid model is the traffic flow database, so the demand of data volume is large. Once the data volume is insufficient, the prediction accuracy will be depressed. Xiao et al. [11] proposed an improved binding cycle truncation accumulated generating operation seasonal grey rolling forecasting model based on the properties of similar seasonality within intraday and weekly trends. The model weakens the random disturbance and highlights the intrinsic grey exponent rule after accumulating the sequence, so that the model has better performance under different traffic flow conditions. Lin-chao et al. [12] proposed a short-term traffic flow prediction model

based on support vector regression which is suitable for real-time monitoring and then analyzed the model parameters by grid search under the premise of considering the influence of space-time factors. The model can achieve ideal prediction effect even if the real-time traffic data are insufficient. Nevertheless, the accuracy of the model will decrease when the traffic status changes. In order to solve such problem, Ma et al. [13] proposed a two-dimensional prediction method by using the Kalman filtering theory based on historical data. The advantage of this method is that the two predicted values are fused by using an equation with weight coefficients where the weight coefficients can be generated in real time in the process of prediction. Experimental results show that the model has an admirable predictive effect. Chan et al. [14] used exponential smoothing to preprocess traffic data that is taken as the input of network and then used Levenberg Marquardt (LM) variant algorithm to train the network weights, making the generalization ability of network enhanced.

However, thanks to the diversified developments of traffic information processing and data transmission techniques within heterogeneous traffic network, as well as the impact caused by dynamic changes of road topology, traffic accidents, severe weather, driving styles, and so forth, short-term traffic data are instantaneous and the irregular volatility is always changing [15, 16]. In this case, the above combinational algorithm can just solve the short-term traffic prediction problem in certain circumstance, and the prediction effect can hardly be achieved in the light of the real-time requirement. Hence, it is urgent to construct new models and algorithms to deal with the short-term traffic forecasting problem aiming at the strong fluctuation triggered by events. Thus, in this paper, we intend to propose a novel method for short-term traffic prediction based on the energetically grey Verhulst prediction algorithm and the first-order difference exponential smoothing technique to solve the problem of low prediction accuracy caused by traffic event-triggering strong fluctuation.

The layout of this paper is arranged as follows. Firstly, based on the introduction of the traditional grey Verhulst model, an improved grey Markov forecasting model is devised by introducing Markov chain. Secondly, combining the advantages of forecasting by utilizing first-order difference exponential smoothing algorithm and introducing a dynamic weighting factor, a new method for short-term traffic forecasting is concretely constructed according to the afore contrived models. Finally, comparative analysis of the examples illustrated that our proposed model and algorithm are more effective.

2. Short-Term Traffic Flow Prediction Model Based on Grey Difference Exponential Smoothing

2.1. Short-Term Traffic Flow Prediction Model Based on Grey Markov Theory

2.1.1. Introduction of Grey Verhulst Model. The grey system theory was first put forward by Professor Deng Julong,

a Chinese scholar, in 1980s [17]. Its quantitative model is mainly based on the structure of number generating, which makes the prediction effect no longer affected by the empirical statistical law gained from the analysis of mass data. At the same time, it surmounted the limitation of white system and the black system relying solely on the probability and statistics method. Therefore, the grey system theory has been widely used in the fields of agricultural production, industrial control, traffic management, and so forth [18].

Classical grey system theory mainly includes GM (1, 1) model, GM (2, 1) model, grey Verhulst model, and so forth [19]. Considering the strong randomness and nonlinearity of traffic flow with strong fluctuation in heterogeneous information model, the data sequence significantly gives expression to the characteristics of nonmonotonic oscillatory development or the S-shaped property of saturation trend, which makes the grey GM (1, 1) prediction model unsuitable. Therefore, we choose the grey Verhulst model in this paper to forecast short-term traffic. The basic modeling process is described as follows.

The nonnegative data sequence is defined as

$$x^{(0)} = (x_1^{(0)}, x_2^{(0)}, \dots, x_n^{(0)}). \quad (1)$$

while $x^{(1)}$ stands for the one-time accumulation sequence (1-Ago):

$$x^{(1)} = (x_1^{(1)}, x_2^{(1)}, \dots, x_n^{(1)}),$$

$$x_k^{(1)} = \sum_{i=1}^k x_i^{(0)} \quad (k = 1, 2, \dots, n). \quad (2)$$

Then, assuming that $z^{(1)}$ represents the consecutive neighbor sequence of $x^{(1)}$,

$$z^{(1)} = (z_1^{(1)}, z_2^{(1)}, \dots, z_n^{(1)}),$$

$$z_k^{(1)} = 0.5(x_k^{(1)} + x_{k-1}^{(1)}) \quad (k = 2, 3, \dots, n). \quad (3)$$

The grey Verhulst model and its whitening equation can be defined as follows.

Definition 1 (see [20, 21]). The grey Verhulst model is

$$x^{(0)} + az^{(1)} = b(z^{(1)})^2, \quad (4)$$

where a and b are the parameters of the equation.

Definition 2 (see [20, 21]). The whitening equation of the grey Verhulst model is

$$\frac{dx^{(1)}}{dt} + ax^{(1)} = b(x^{(1)})^2, \quad (5)$$

where a , b , and t are the parameters of the whitening equation.

As deduced in [20, 21], the following conclusions are drawn.

Theorem 3. If the grey Verhulst model is defined as formula (4) and $\hat{a} = (a, b)^T$ represents the parameter column, let

$$B = \begin{bmatrix} -z_2^{(1)} & (z_2^{(1)})^2 \\ -z_3^{(1)} & (z_3^{(1)})^2 \\ \vdots & \vdots \\ -z_n^{(1)} & (z_n^{(1)})^2 \end{bmatrix}, \quad (6)$$

$$Y = \begin{bmatrix} x_2^{(0)} \\ x_3^{(0)} \\ \vdots \\ x_n^{(0)} \end{bmatrix}.$$

The least squares estimator of the parameter column satisfies

$$\hat{\alpha} = (B^T B)^{-1} B^T Y. \quad (7)$$

Theorem 4. If the grey Verhulst model is defined as definition (2), the solution of the whitening equation can be deduced as

$$x^{(1)}(t) = \frac{ax_0^{(1)}}{bx_0^{(1)} + (a - bx_0^{(1)})e^{at}}. \quad (8)$$

According to Theorems 3 and 4, the following conclusions are drawn.

Inference 5. The time response sequence of grey Verhulst model can be defined as

$$\hat{x}_{k+1}^{(1)} = \frac{ax_0^{(1)}}{bx_0^{(1)} + (a - b_0^{(1)})e^{at}}, \quad (k = 1, 2, \dots, n). \quad (9)$$

Let $x_0^{(1)} = x_1^{(0)}$; we iterate formula (9) to be

$$\hat{x}_{k+1}^{(1)} = \frac{ax_1^{(0)}}{bx_1^{(0)} + (a - bx_1^{(0)})e^{at}}, \quad (k = 1, 2, \dots, n). \quad (10)$$

The reduction formula can then be defined as

$$\hat{x}_{k+1}^{(0)} = \hat{x}_{k+1}^{(1)} - \hat{x}_k^{(1)}, \quad (k = 1, 2, \dots, n). \quad (11)$$

Pointing at the traffic flow with the trend of increasing saturation, numerous traffic prediction algorithms have been proposed taking advantage of the grey Verhulst model to achieve preferable prediction effects. However, as a complicated nonlinear system involving multitudinous uncertainties, the probability of fortuitous events on urban roads is highly fluctuant, which leads to the deviation of forecasting results. Therefore, it is necessary to take the nonlinearity and time-varying characteristics of overall interactions into account in line with various influential factors. So, we improve the algorithm aiming at the accuracy of prediction model next.

2.1.2. Short-Term Traffic Flow Prediction Model Based on Grey Markov Theory. During the process of traffic data aggregation, which is coordinated by complex human-vehicle-environment interaction, the current traffic flow is often affected by previous moments. Therefore, in order to improve the prediction accuracy of grey Verhulst model, mathematical description of traffic flow aggregation is given in advance by constructing the Markov state transition probability matrix [22].

It is obvious that the state transition probability matrix should be updated over time when Markov is used to optimize the grey model. That is to say, at time point $t + 1$, the state transition probability of Markov chain should be recalculated according to real-time traffic data along with the state transition probability matrix updated synchronously. Therefore, at time point $t + 1$, the probability of traffic flow aggregation from state i to j can be described as

$$P_{ij} = p(X_{t+1} = j | X_t = i). \quad (12)$$

And the updated Markov transition probability matrix is formulated to be

$$P = \begin{bmatrix} p_{11} & p_{12} & \cdots & p_{1n} \\ p_{21} & p_{22} & \cdots & p_{2n} \\ \vdots & \vdots & & \vdots \\ p_{n1} & p_{n2} & \cdots & p_{nn} \end{bmatrix}, \quad (13)$$

where each element p_{ij} is constrained as $p_{ij} \geq 0, i, j \in I$, and $\sum_{j \in I} p_{ij} = 1, i \in I$.

The grey Markov forecasting model is shown in Figure 1. According to the flow chart of this model, the value of grey Verhulst model to be predicted is \hat{x} at time point $t + 1$. In order to make the predicted value as close to the true value as possible, the true value at the time point t is utilized and represented as x_{ij} , and the transition probability from the true value of time point t to the predictive value is denoted by p_{ij} . Thus, the cumulative sum of the product of the traffic flow average value and its corresponding probability from the current state to other states can be taken as the predicted value of the grey Markov model:

$$G_{t+1} = \begin{cases} \hat{x} + \phi & (p_{ij} = 0) \\ \hat{x} \cdot p_{ij} + \phi + \sum_{k=1, k \neq j}^n p_{ik} \cdot \bar{x}_{ik} & (p_{ij} \neq 0), \end{cases} \quad (14)$$

where ϕ is the adjustment factor and $\phi = \hat{x}_t - \hat{x}_{t-1}$.

The adjustment factor ϕ is used to solve the data fluctuation problem of collaborative traffic flow prediction in terms of human, vehicle, and environment interaction. However, the smoothness of short-term traffic flow forecasting remains to be improved once events trigger strong fluctuations caused by traffic flow aggregation. Therefore, in order to continuously improve the prediction smoothness together with accuracy, a grey difference exponential smoothing short-term traffic prediction method is further constructed by introducing the first-order difference exponential smoothing model in the next subsection.

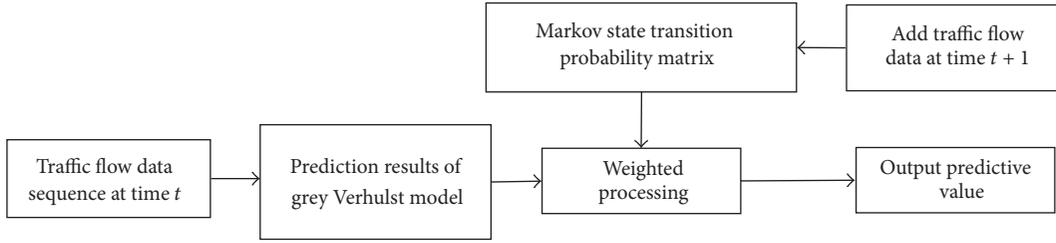


FIGURE 1: Grey forecasting model based on Markov chain.

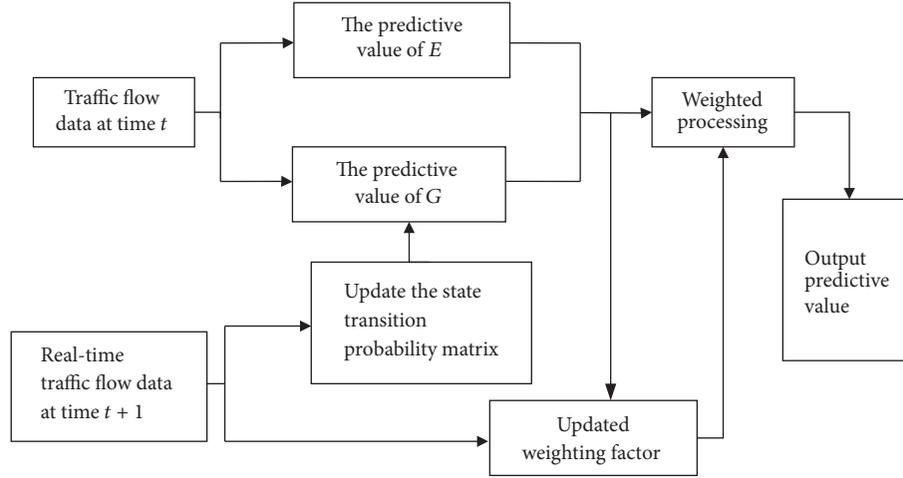


FIGURE 2: Grey difference exponential smoothing forecasting model.

2.2. Short-Term Flow Traffic Prediction Model Based on Grey Difference Exponential Smoothing. Exponential smoothing [23] is a widely used sequential prediction algorithm, whose predictive values are obtained by assigning different weights to historical records. Similar to the exponential smoothing algorithm, the first-order difference exponential smoothing algorithm achieves the predicted value by giving different weights to different data sequences. The basic modeling process is shown as follows.

For data sequence (1), the difference sequence is represented as

$$\nabla \hat{E}_t = x_{t+1}^{(0)} - x_t^{(0)}, \quad (t = 1, 2, \dots, n-1). \quad (15)$$

And the first-order difference exponential smoothing equation is

$$E_{t+1} = \nabla E_t + x_t^{(0)}, \quad (t = 1, 2, \dots, n), \quad (16)$$

where $\nabla E_t = \alpha \nabla \hat{E}_{t-1} + (1-\alpha) \nabla E_{t-1}$. Therefore, the forecast result should be

$$E_{t+1} = \alpha \sum_{j=1}^{t-1} (1-\alpha)^{j-1} \nabla \hat{E}_{t-j} + (1-\alpha)^{t-1} \nabla E_1 + x_t^{(0)}. \quad (17)$$

Herein, $0 < \alpha < 1$ and $(1-\alpha)^{t-1} \nabla E_1 \rightarrow 0$ when $t \rightarrow \infty$. Thus, formula (17) can be rewritten as

$$E_{t+1} = \alpha \sum_{j=1}^{t-1} (1-\alpha)^{j-1} \nabla \hat{E}_{t-j} + x_t^{(0)}. \quad (18)$$

We use dynamic weighting factors to weight the models of grey Markov model and first-order difference exponential smoothing model. The predictive value of the grey exponential smoothing model is thus achieved:

$$Y_{t+1} = \varphi G_{t+1} + \gamma E_{t+1}, \quad (\varphi + \gamma = 1). \quad (19)$$

In formula (19), G_{t+1} and E_{t+1} are the predictive value of grey Markov and first-order difference exponential smoothing, respectively, while φ and γ are their corresponding weight factors. The weight factors are calculated as follows:

$$\varphi = \frac{|x_{t+1}^{(0)} - E_{t+1}| + \zeta/2}{|x_{t+1}^{(0)} - G_{t+1}| + |x_{t+1}^{(0)} - E_{t+1}| + \zeta}, \quad (20)$$

$$\gamma = \frac{|x_{t+1}^{(0)} - G_{t+1}| + \zeta/2}{|x_{t+1}^{(0)} - G_{t+1}| + |x_{t+1}^{(0)} - E_{t+1}| + \zeta}.$$

In formula (20), the weight of φ and γ depends on the absolute value of the difference between the predicted results of different models and the short-term traffic flow. In the meantime, in order to avoid the failure of formula (20), when the two prediction errors are 0, ζ tends to zero. The pseudocode of grey difference exponential smoothing prediction model is shown in Algorithm 1.

According to the synthetic forecasting algorithm, the prediction process of grey difference exponential smoothing prediction model is shown in Figure 2.

Input: Historical traffic flow data set $x^{(0)}$
Output: Short-term traffic flow forecasting result set Y

- (1) **Initial state probability matrix** $P, \alpha, \varphi, \gamma, \xi$
- (2) **for** $t \leftarrow 1$ to n **do**
- (3) **Calculate forecasting result** E_{t+1} **and** G_{t+1}
- (4) $Y_{t+1} = \varphi G_{t+1} + \gamma E_{t+1}, (\varphi + \gamma = 1)$
- (5) **Calculate the error** ΔE_{t+1} **and** ΔG_{t+1}
- (6) $\Delta E_{t+1} = |x^{(0)}_{t+1} - E_{t+1}|, \Delta G_{t+1} = |x^{(0)}_{t+1} - G_{t+1}|$
- (7) **Update the weights of** φ **and** γ
- (8) $\varphi = \frac{\Delta E_{t+1} + \xi/2}{\Delta G_{t+1} + \Delta E_{t+1} + \xi}, \gamma = \frac{\Delta G_{t+1} + \xi/2}{\Delta G_{t+1} + \Delta E_{t+1} + \xi}$
- (9) **Update the state transition matrix** P **by** X_{t+1}
- (10) **end**
- (11) **return** Y

ALGORITHM 1: Pseudocode of grey differential exponent smoothing.

3. Example and Analysis

In order to verify the prediction effect of our model, the benchmarks of mean absolute error (MAE), mean square error (MSE), mean absolute percentage error (MAPE), and mean square percentage error (MSPE) are taken as evaluation indexes. Those evaluation indexes are defined as formulas (21)–(24). In those formulas, X_t and \hat{Y}_t are the real value and predicted value of the traffic flow at time t , respectively, and n stands for the number of samples in the short-term traffic flow.

(1) Mean absolute error is

$$\text{MAE} = \frac{1}{n} \sum_{t=1}^n |X_t - \hat{Y}_t|. \quad (21)$$

(2) Mean square error is

$$\text{MSE} = \frac{1}{n} \sqrt{\sum_{t=1}^n (X_t - \hat{Y}_t)^2}. \quad (22)$$

(3) Mean absolute percentage error is

$$\text{MAPE} = \frac{1}{n} \sum_{t=1}^n \left| \frac{X_t - \hat{Y}_t}{X_t} \right|. \quad (23)$$

(4) Mean square percentage error is

$$\text{MSPE} = \frac{1}{n} \sqrt{\sum_{t=1}^n \left(\frac{X_t - \hat{Y}_t}{X_t} \right)^2}. \quad (24)$$

The urban expressway in Nan'an District of Chongqing Municipality in China is selected in observation phase to verify the actual prediction effect of the grey difference exponential smoothing model, which is composed of 6 bidirectional lanes. The schematic diagram of the section observation is shown in Figure 3. The data detected include the

TABLE 1: Evaluation index of first-order difference exponential smoothing algorithm.

| α | Index | | | |
|-----------------|---------------|--------------|--------------|--------------|
| | MAE | MSE | MAPE | MSPE |
| $\alpha = 0.05$ | 26.820 | 2.563 | 0.389 | 0.046 |
| $\alpha = 0.08$ | 27.191 | 2.601 | 0.392 | 0.046 |
| $\alpha = 0.10$ | 27.451 | 2.627 | 0.394 | 0.046 |
| $\alpha = 0.20$ | 28.828 | 2.760 | 0.408 | 0.047 |
| $\alpha = 0.50$ | 33.704 | 3.220 | 0.470 | 0.052 |

traffic volume, speed, and occupancy of each lane, which are collected via microwave-based vehicle sensors.

The road traffic flow is counted every 5 minutes, and the 226 observation data points of No. 1 Lane are used as the basic data for verification. The fluctuation trend of the data is shown in Figure 4.

As illustrated in Figure 4, there is a strong fluctuation in historical traffic flow, which brought about great difficulties to the prediction. When using the first-order difference exponential smoothing algorithm to predict the short time traffic flow, it is necessary to determine the value of α in the definite formula (18). In this paper, five experiments are carried out according to the parameters listed in Table 1, and the evaluation indexes corresponding to the first-order difference exponential smoothing algorithm are obtained in terms of different benchmarks.

According to the results in Table 1, when $\alpha = 0.05$, the first-order difference exponential smoothing algorithm produces the best effect, and the fitting patterns between the predicted value and actual value are shown in Figure 5.

The initial state probability transfer matrix is calculated from the historical traffic sequence when the grey Markov model is used to predict the short-term traffic flow. The state is divided according to the same interval that every 30 vehicles are used as a state. In the following prediction process, we added real-time traffic data to update Markov state probability transfer matrix in real time. The prediction result of grey Markov model is determined by formula (14), and the fitting effect of predicted value and actual value is shown in Figure 6.

Comparing Figure 6 with Figure 5, the grey Markov model can commendably reflect the trend of traffic flow. However, urban road traffic as a complex nonlinear system shows high uncertainty at different times. The grey Markov chain model cannot accurately predict once sharp fluctuations occur. Meanwhile, compared with the grey Markov chain prediction model, the first-order difference exponential smoothing model is more accordant with the trend of original data, but the prediction results of the algorithm depend on the value of α and, more importantly, are affected by time delays.

Therefore, combining the advantages of grey Markov model and first-order difference exponential smoothing, a new predictive value is obtained by using dynamic weighting factors according to formula (19) and formula (20). The prediction results are shown in Figure 7.

Based on the short-term traffic data of the observed road sections, the test data are predicted, respectively, by formulas

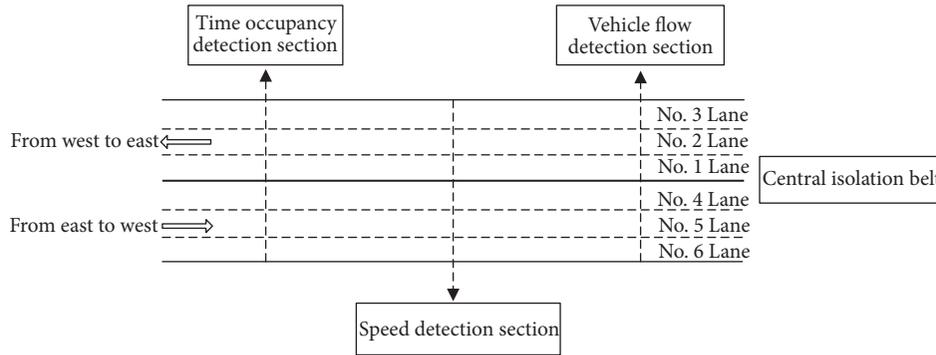


FIGURE 3: Road observation diagram.

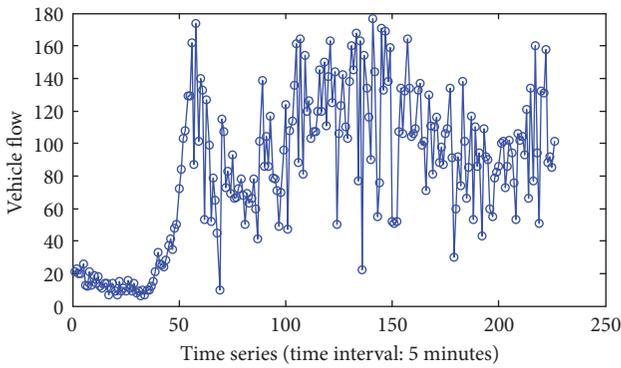
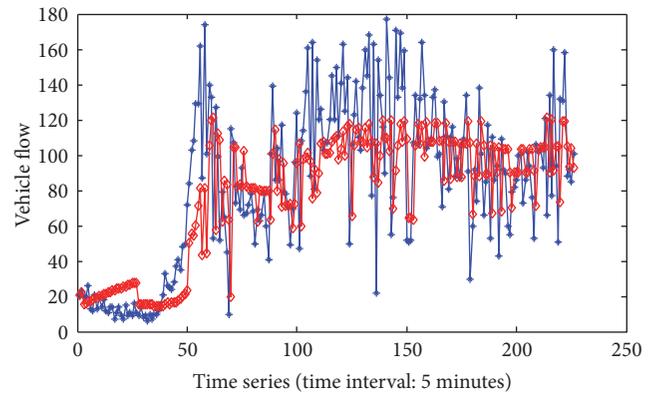
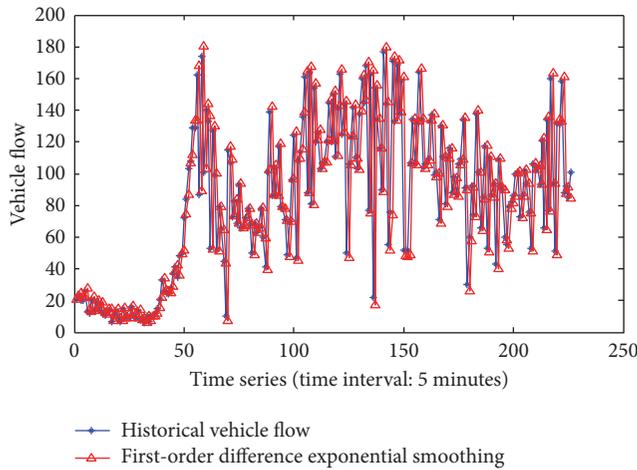


FIGURE 4: Historical traffic flow data.



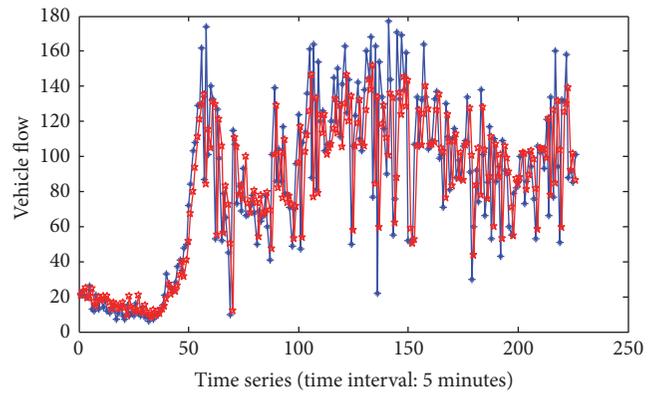
— Historical vehicle flow
— Grey Markov model

FIGURE 6: Predictive results of grey Markov model.



— Historical vehicle flow
— First-order difference exponential smoothing

FIGURE 5: Predictive results of first-order difference exponential smoothing model.



— Historical vehicle flow
— Grey difference exponential smoothing

FIGURE 7: Predictive results of grey difference exponential smoothing model.

(14), (18), and (19). The fitting results of the first 120 data points and the last 106 data points are separately illustrated in Figures 8 and 9.

The evaluation indexes of the above prediction models are shown in Table 2 for the first-order difference exponential smoothing parameter $\alpha = 0.05$. Simulation results showed that the grey difference exponential smoothing algorithm can fit short-term traffic flow well and can be applied to urban

traffic signal control for short-term traffic flow data with strong fluctuation.

Compared with the above models, the first-order difference exponential smoothing algorithm can closely reflect the

TABLE 2: Evaluation index of forecasting model.

| Model | Index | | | |
|--|---------------|--------------|--------------|--------------|
| | MAE | MSE | MAPE | MSPE |
| Grey Markov model | 25.520 | 2.220 | 0.445 | 0.049 |
| First-order difference exponential smoothing model | 26.820 | 2.563 | 0.389 | 0.046 |
| Grey difference exponential smoothing model | 23.943 | 2.237 | 0.368 | 0.043 |

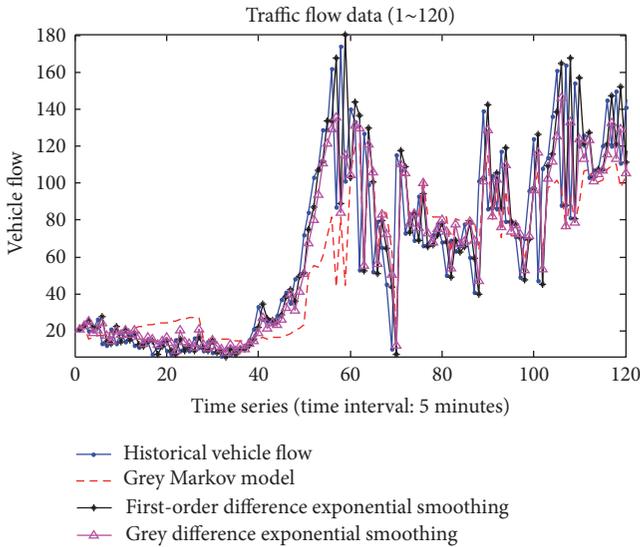


FIGURE 8: Fitting curves between model predictive values and actual test values (the first 120 data points).

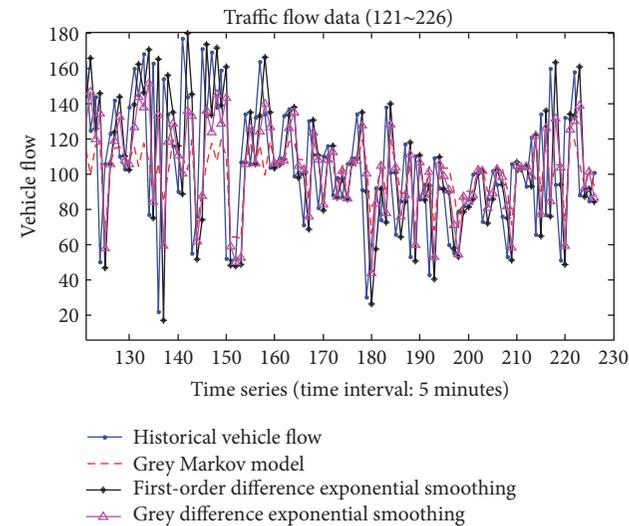


FIGURE 9: Fitting curves between model predictive values and actual test values (the latter 106 data points).

trend of the original data sequence in the prediction trend of the data, but its predicted value depends on the value of α and the trend of the data sequence. For the abrupt data points, the accuracy of its prediction will often be affected to distinct degrees. Compared with the first-order difference exponential smoothing model, the prediction accuracy of grey Markov model considers the development trend of the

whole data and the state transition probability of historical traffic flow data. Therefore, the model can reflect the overall equilibrium level for data series with large fluctuation. The grey difference exponential smoothing model combines the advantages of the two models and uses the dynamic weight to obtain the predictive value. The performance evaluation indexes MAE, MSE, MAPE, and MSPE of the improved method are better than the first-order difference exponential smoothing model and grey Markov model for the same traffic flow sequence. However, it can also be seen from Figure 6 and evaluation indexes that the prediction effect of the grey difference exponential smoothing algorithm using dynamic weighted calculation is not absolutely well. Therefore, the method needs to be further optimized and improved.

4. Conclusion

This paper mainly contributes to the prediction effect of short-term traffic flow. To overcome the shortcoming of low accuracy in short-term traffic flow prediction caused by strong traffic flow fluctuation, a novel method for short-term traffic forecasting based on the combination of improved grey Verhulst prediction algorithm and first-order difference exponential smoothing is proposed. The main conclusions are as follows.

(1) The grey differential exponential smoothing model combines the advantages of the grey Markov model and the first-order differential exponential smoothing model. The experimental results illustrated that this method is suitable for forecasting short-term traffic flow with large fluctuation. Therefore, this method is practical and feasible.

The analysis shows that this method has smaller prediction error compared with the grey Markov model and the first-order difference exponential smoothing model, and the prediction result is closer to the actual value.

(2) In a word, the method proposed in this paper has obvious advantages. However, due to high uncertainty and nonlinearity of the short-term traffic flow, how to improve the prediction accuracy with fluctuation remains an open problem. So this will be our continuous research orientation.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work is supported by the National Science Foundation of China (NSFC) under Grants 61573076 and 61703063; the Scientific Research Foundation for the Returned Overseas Chinese Scholars under Grant 2015-49; the Program for Excellent

Talents of Chongqing Higher School under Grant 2014-18; Science and Technology Research Project of Chongqing Municipal Education Commission of China under Grants KJ1705121 and KJ1705139; the Chongqing Natural Science Foundation of China under Grant CSTC2017jcyjA1665.

References

- [1] Z. Hong-bin, S. Xiao-duan, and H. Yu-long, "Analysis and prediction of complex dynamical characteristics of short-term traffic flow," *Acta Physica Sinica*, vol. 63, no. 4, Article ID 080502, pp. 51–58, 2014 (Chinese).
- [2] D.-R. Huang, J. Song, S.-Q. Li, and H.-Y. Xiang, "Control technology review of traffic congestion in urban road network under networked dynamic scheduling and control," *Journal of Traffic and Transportation Engineering*, vol. 13, no. 5, pp. 105–114, 2013 (Chinese).
- [3] S. Oh, Y.-J. Byon, K. Jang, and H. Yeo, "Short-term travel-time prediction on highway: a review on model-based approach," *KSCSE Journal of Civil Engineering*, vol. 3, pp. 1–13, 2017.
- [4] N. Fan, X.-M. Zhao, M. Dai, and Y.-S. An, "Short-term traffic flow prediction model," *Journal of Traffic and Transportation Engineering*, vol. 12, no. 4, pp. 114–119, 2012 (Chinese).
- [5] Y.-A. Daraghmi, C.-W. Yi, and T.-C. Chiang, "Negative binomial additive models for short-term traffic flow forecasting in Urban areas," *IEEE Transactions on Intelligent Transportation Systems*, vol. 15, no. 2, pp. 784–793, 2014.
- [6] Y.-S. Jeong, Y.-J. Byon, M. M. Castro-Neto, and S. M. Easa, "Supervised weighting-online learning algorithm for short-term traffic flow prediction," *IEEE Transactions on Intelligent Transportation Systems*, vol. 14, no. 4, pp. 1700–1707, 2013.
- [7] K. Y. Chan, T. Dillon, E. Chang, and J. Singh, "Prediction of short-term traffic variables using intelligent swarm-based neural networks," *IEEE Transactions on Control Systems Technology*, vol. 21, no. 1, pp. 263–274, 2013.
- [8] M. Lippi, M. Bertini, and P. Frasconi, "Short-term traffic flow forecasting: an experimental comparison of time-series analysis and supervised learning," *IEEE Transactions on Intelligent Transportation Systems*, vol. 14, no. 2, pp. 871–882, 2013.
- [9] N. G. Polson and V. O. Sokolov, "Deep learning for short-term traffic flow prediction," *Transportation Research Part C: Emerging Technologies*, vol. 79, pp. 1–17, 2017.
- [10] H.-H. Xie, X.-H. Dai, and Y. Qi, "Improved K-nearest neighbor algorithm for short-term traffic flow forecasting," *Journal of Traffic and Transportation Engineering*, vol. 14, no. 3, pp. 87–94, 2014 (Chinese).
- [11] X. Xiao, J. Yang, S. Mao, and J. Wen, "An improved seasonal rolling grey forecasting model using a cycle truncation accumulated generating operation for traffic flow," *Applied Mathematical Modelling*, vol. 51, pp. 386–404, 2017.
- [12] L. Lin-chao, H. Shang-lu, and Z. Jian, "Online short-term traffic flow prediction considering the impact of temporal-spatial features," *Journal of Transportation Systems Engineering and Information Technology*, vol. 16, no. 5, pp. 165–171, 2016 (Chinese).
- [13] M. Ma, S. Liang, H. Guo, and J. Yang, "Short-term traffic flow prediction using a self-adaptive two-dimensional forecasting method," *Advances in Mechanical Engineering*, vol. 9, no. 8, pp. 1–12, 2017.
- [14] K. Y. Chan, T. S. Dillon, J. Singh, and E. Chang, "Neural-network-based models for short-term traffic flow forecasting using a hybrid exponential smoothing and levenberg-marquardt algorithm," *IEEE Transactions on Intelligent Transportation Systems*, vol. 13, no. 2, pp. 644–654, 2012.
- [15] X. Wang, A. J. Khattak, J. Liu, G. Masghati-Amoli, and S. Son, "What is the level of volatility in instantaneous driving decisions?" *Transportation Research Part C: Emerging Technologies*, vol. 58, pp. 413–427, 2015.
- [16] W. Qing-bo, H. Zhao-cheng, and Z. Xi-shuang, "Prediction of Urban Traffic Performance Index Considering Multiple Factors," *Journal of Transportation Systems Engineering and Information Technology*, vol. 17, no. 1, pp. 74–81, 2017 (Chinese).
- [17] S. Liu, Y. Yang, N. Xie et al., "New progress of grey system theory in the new millennium," *The Journal of Grey System*, vol. 6, no. 1, pp. 2–31, 2016.
- [18] R. Rajesh, V. Ravi, and R. Venkata Rao, "Selection of risk mitigation strategy in electronic supply chains using grey theory and digraph-matrix approaches," *International Journal of Production Research*, vol. 53, no. 1, pp. 238–257, 2015.
- [19] H. Da-rong and H. Li-fen, "Present situation and development of grey system theory in fault forecast application," *Journal of Gun Launch & Control*, vol. 3, pp. 88–92, 2009 (Chinese).
- [20] D. Song, D. Yao-guo, X. Ning et al., "Optimized background value in grey Verhulst model and its application," *Control and Decision*, vol. 30, no. 10, pp. 1835–1840, 2015 (Chinese).
- [21] Y. Zhang, "Improved background value of grey Verhulst model and its application," *System Engineering-Theory and Practice*, vol. 33, no. 12, pp. 3168–3171, 2013.
- [22] Z. Yi-fan and C. Wen-ying, "Markov chain analysis of metro network in responding to emergency," *China Safety Science Journal*, vol. 25, no. 2, pp. 165–170, 2015 (Chinese).
- [23] Q. Chi and H. Zhong-sheng, "Application of adaptive sing-exponent smoothing for short-term traffic flow prediction," *Control Theory & Applications*, vol. 29, no. 4, pp. 465–469, 2012 (Chinese).

Research Article

Detection of Two-Level Inverter Open-Circuit Fault Using a Combined DWT-NN Approach

Bilal Djamal Eddine Cherif  and **Azeddine Bendiabdellah**

Diagnostic Group, Laboratory LDEE, Electrical Engineering Faculty, University of Sciences and Technology of Oran, Bir El Djir, Algeria

Correspondence should be addressed to Bilal Djamal Eddine Cherif; cherif.doc84@gmail.com

Received 18 November 2017; Revised 19 January 2018; Accepted 29 January 2018; Published 11 March 2018

Academic Editor: Qiang Liu

Copyright © 2018 Bilal Djamal Eddine Cherif and Azeddine Bendiabdellah. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Three-phase static converters with voltage structure are widely used in many industrial systems. In order to prevent the propagation of the fault to other components of the system and ensure continuity of service in the event of a failure of the converter, efficient and rapid methods of detection and localization must be implemented. This paper work addresses a diagnostic technique based on the discrete wavelet transform (DWT) algorithm and the approach of neural network (NN), for the detection of an inverter IGBT open-circuit switch fault. To illustrate the merits of the technique and validate the results, experimental tests are conducted using a built voltage inverter fed induction motor. The inverter is controlled by the SVM control strategy.

1. Introduction

Several researchers have carried out their investigation in relation to the field of detection and location of faults in static converters and more particularly those related to three-phase power inverters [1]. The treated fault is mainly concerned with the open-circuit fault of an inverter IGBT switch [2]. Most published papers are based on Park's current vectors approach [3]. This approach is based on the trajectory tracking of the phase current vector. In fact, for the case of a healthy state condition of the inverter, the trajectory of these current vectors in the $(d-q)$ frame is a circle. It was found that the circle becomes a semicircle under an open-circuit IGBT switch fault in one of the legs of the inverter. The position of this semicircle in the $(d-q)$ frame makes it possible to identify the faulty IGBT switch [4]. Another paper used the mean value of the phase currents in Park's frame for the extraction of the open-circuit fault angle of each IGBT switch [1–5]; unfortunately this method presents an inconvenience as it depends on the load. To overcome the problem, some authors suggested the normalized DC current method which is fundamentally based on the DC component of the current and the first-order harmonic coefficients of the alternating current (AC) [6]. Some detection techniques mentioned above are briefly discussed in [7, 8].

Others researchers used stator current as key parameter for fault diagnosis purpose because it does not require costly sensors. This technique is widely termed as Motor Current Signature Analysis (MCSA) technique [9–11]. For steady state constant load conditions, the Fast Fourier Transform (FFT) algorithm has been used for different induction motor fault diagnosis purpose. The (FFT) algorithm is able to diagnose bearing fault for full load conditions but not for the no-load or light load condition [7]. Therefore, some researchers proposed short-term Fourier transform (STFT) and wavelet transform (WT) methods for different fault diagnosis purpose of the induction motor by taking into consideration the quantity of time (that is for nonstationary applications). However, the (STFT) method faces a big disadvantage; it gives poor frequency resolution because it shows constant window size for all frequencies.

The wavelet transform (WT) technique has perhaps been the most persistent development in recent decades. It has drawn the attention of several researchers in various fields, such as signal processing, image processing, communication, computer science, and mathematics. Numerous works describing the advancement in wavelet theory and its applications in various fields have been published. Among others, it is one of the most attractive techniques in the field of

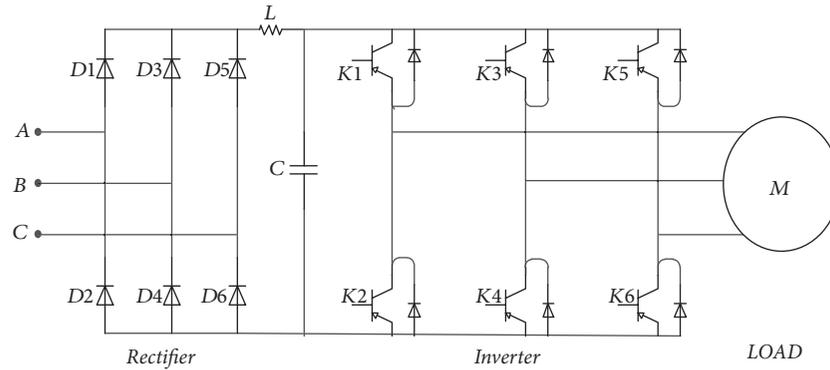


FIGURE 1: Structure of a three-phase two-level inverter.

rotary machines and static converter for fault diagnosis and it particularly matches well for the nonstationary signals [12, 13]. The main disadvantage of this technique is that the length of the range (scale) is fixed. Therefore, it is necessary to be able to keep the time-frequency representation and carry out an analysis based on a concept somewhat different from the concept of frequency: the scale concept [14]. In 1982, J. Morlet opens the way to the solution by constructing the wavelet analysis, based on the concept of scale. By manipulating the scale factor, one can zoom in and out a portion of the signal [14, 15]. Unlike the short-term Fourier transform, the wavelet transform uses the notion of time-scale involving dynamic length analysis windows. Because the continuous wavelet transform is very sensitive to noise, the discrete wavelet transform is preferred and used in our paper.

The application of artificial intelligence (AI) based techniques can be advantageous in fault diagnosis since these diagnoses have several advantages. For instance, since AI-based techniques do not require mathematical models, the development time can be significantly reduced. A literature review of recent developments in the field of AI-based diagnostic systems in power inverters has been presented [16]. In addition, some authors have studied the application of a neural network (NN) to establish a fault diagnosis system and judge the faults of power transistors [17]. These NN-based fault diagnosis methods allow an accurate solution to a particular fault problem without accurate knowledge of the faulty system. However, their main drawback is due to the fact that the exact architecture of the NN to be used is generally not known in advance.

The work proposed in this paper addresses an open-circuit fault detection of an IGBT switch of an inverter controlled by a DSPACE 1104 card based on the SVM control strategy feeding an induction motor. The analysis tools and fault diagnosis are based on the use of a combined DWT-NN approach. The DWT algorithm focuses on the investigation of the details of the stator current signal. The variation in these details for both healthy and faulty inverter cases enables us to extract useful information related to the open-circuit inverter switch faults. The NN approach is introduced in order to automate the fault diagnostic system by including a learning phase that helps in developing a very rich database storing relevant information about the open-circuit faults. To

assess the effectiveness and merits of the proposed approach and validate the obtained results, experimental tests are conducted by the group diagnostic at the LDEE laboratory.

2. Voltage Source Inverter Structure

Figure 1 shows the structure of a three-phase two-level voltage source inverter feeding an induction motor.

This inverter is controlled by the SVM control strategy. For each leg of the inverter, there are two possible states:

- (i) *State 1*: the higher switch K_X ($X = 1, 2$ or 3) is closed, while the lower switch K_X ($X = 1, 2$ or 3) is open. The output voltage relative to the neutral of the DC source is V_{dc} .
- (ii) *State 0*: the lower switch K_X ($X = 1, 2$ or 3) is closed, while the higher switch K_X ($X = 1, 2$ or 3) is open. The output voltage relative to the neutral of the DC source is 0 v.

Unfortunately during its operation, various failures can affect the inverter especially in terms of its so-called power components (IGBT in our case study) because of their fragility. Two types of faults can be reported [1]:

- (i) Short-circuit faults affecting the IGBT switches are the most serious faults. In the presence of such a fault, the current reaches limits which can cause the fusion of its chip or its connection. If the detection of this type of fault does not occur rapidly (less than 10 microseconds), then the IGBT switch which is still active on the same leg undergoes the same phenomenon and so the whole inverter leg is shorted.
- (ii) Open-circuit faults affecting the IGBT switches may occur when, for any reason, the IGBT is disconnected, is damaged, or had a problem in its grid control signal. This type of fault is very difficult to perceive directly because the motor can continue to operate but with a degradation of its performance due to the occurrence of fluctuations in the mechanical parameters (speed and torque) as well as an imbalance of the currents where the currents of the other two healthy legs take high values to maintain the average torque and the

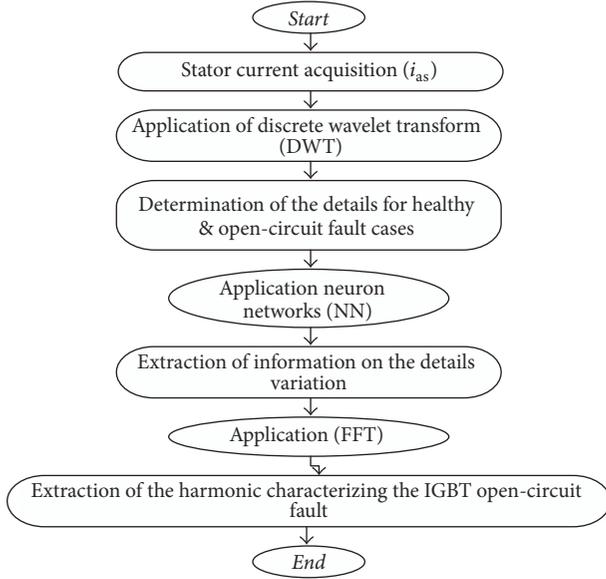


FIGURE 2: Flowchart of the proposed DWT-NN approach.

speed. The starting of the motor in the presence of this type of fault cannot always be ensured; everything will depend on the value of the torque which can be close to zero for certain rotor position.

3. Theoretical Bases of the Combined DWT-NN Approach

The flowchart presented in Figure 2 is to illustrate the main steps required when using the combined DWT-NN approach.

The following two subsections tackle in more detail both the DWT algorithm and the NN approach.

3.1. DWT Algorithm Principle. The wavelet transform is a mathematical tool which allows the decomposition of a temporal signal into a series of coefficients called approximation and detail; the approximations represent the slow variations of the signal, whereas the details represent the fastest [6, 7].

A wavelet $\psi(t)$ is an oscillating function of zero mean, with a certain degree of regularity and whose support is finite (which explains the word “wavelet,” which means small wave). Mathematically, a wavelet $\psi(t)$ is a zero mean function that is expanded by a scale parameter and is translated by the time parameter b “translation in time parameter”; it is defined as follows:

$$\psi_{a,b}(t) = \frac{1}{\sqrt{a}} \Psi\left(\frac{t-b}{a}\right). \quad (1)$$

The expansions and translations of this “mother” wavelet are used to allow the projection of the signal over different time scales. These scales depend mainly on the sampling frequency F_e of the signal to be processed.

Under these conditions, a signal $x(t)$ can be represented as wavelets given by the following equation:

$$x(t) = \sum_{k=-\infty}^{\infty} c_k \varphi(t-k) + \sum_{k=-\infty}^{\infty} \sum_{j=0}^{\infty} d_{j,k} 2^{j/2} \psi(2^j t - k). \quad (2)$$

$\psi(t)$ is the derivative of the function $\varphi(t)$ with

$$c_k = \int_{-\infty}^{\infty} x(t) \varphi(t-k) dt, \quad (3)$$

$$d_{j,k} = \int_{-\infty}^{\infty} x(t) 2^{j/2} \psi(2^j t - k) dt,$$

where c_k are called the approximation coefficients and $d_{j,k}$ the detail coefficients.

In the case of a discrete wavelet transform (DWT) which is the tool of our paper work, the expansion and translation parameters “ a ” and “ b ” of $\psi(t)$ are limited only to discrete values leading to the following expression:

$$\psi_{m,n}(t) = \frac{1}{\sqrt{a_0^m}} \psi\left(\frac{t - nb_0 a_0^m}{a_0^m}\right), \quad (4)$$

where the parameters “ m ” and “ n ” are integers allowing the control of the dilation and the translation of the mother wavelets

Obviously, according to the last equation, different wavelets (Haar, Daubechies, Coiflet, Meyer, Morlet, etc.) generate different wavelet classes and consequently the behavior of the decomposed signal can be very different. In fact, each wavelet has particular characteristics whose choice depends on the desired application. In this paper, the Coiflet wavelets are to be applied for the detection of the IGBT open-circuit. Figure 3 depicts a typical Coiflet wavelet family graphical representation.

3.1.1. Number of Decomposition Levels Required. The DWT using the Coiflet algorithm is based on signal the decomposition using low-pass filters (LPF) and high-pass filters (HPF) followed by subsampling. Then, in each level of decomposition, the coefficients of approximation and details are computed. Figure 4 shows the implementation of the DWT for three decomposition levels as an example.

Where the LPF frequency band is defined as

$$\left[0, \frac{F_e}{2^n}\right], \quad n = 1, 2, \dots, N_{\text{levels}}. \quad (5)$$

And the HPF frequency band is defined as

$$\left[\frac{F_e}{2^n}, \frac{F_e}{2^{(n-1)}}\right], \quad n = 1, 2, \dots, N_{\text{levels}}, \quad (6)$$

where n indicates the number of decomposition levels and the N_{levels} are the maximum number of decomposition levels.

Prior knowledge of N_{levels} of the signal to be processed is essential for a reliable and fast analysis by the Coiflet mother wavelet. The following equation gives this required parameter [11]:

$$N_{\text{levels}} = \text{int}\left(\frac{\log(F_e/f_s)}{\log(2)}\right) + 2, \quad (7)$$

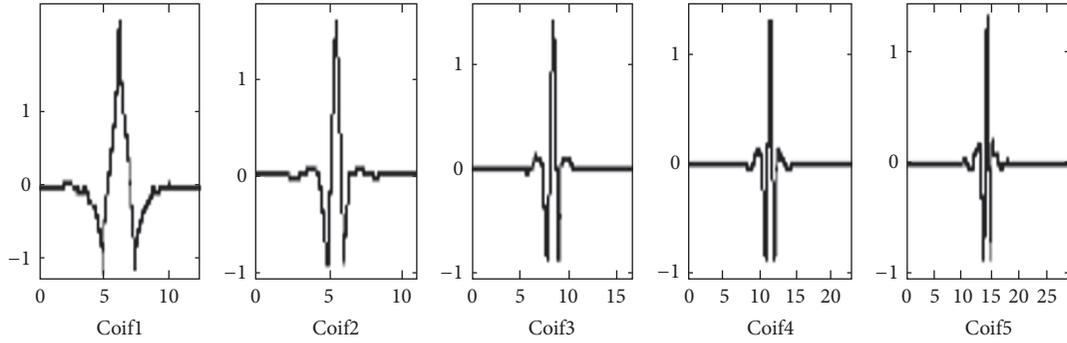


FIGURE 3: Coiflet wavelet families [16].

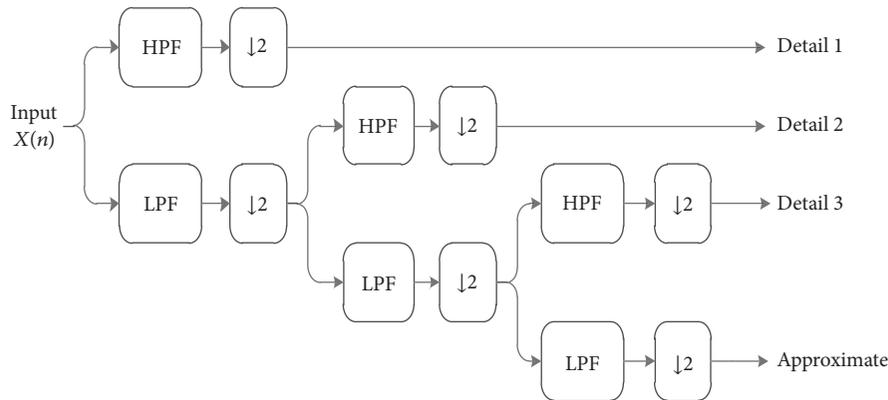


FIGURE 4: DWT implementation procedure.

where f_s is the supply frequency and F_e is the sampling frequency. Note that N_{levels} must obviously be an integer.

Knowing f_s and F_e , one can calculate the number of appropriate decompositions. For our case paper study, considering a sampling frequency of 1500 Hz and a supply frequency of 50 Hz, the number of decomposition levels required is

$$N_{levels} = \text{int} \left(\frac{\log(1500/50)}{\log(2)} \right) + 2 = \text{int}(4.9068) + 2 = 6 \text{ levels.} \tag{8}$$

Using (5) and (6) and the number of levels computed from (8), the decomposition of the motor stator current signal as a function of the sampling frequency for the six levels of decomposition can thus be obtained as illustrated in Figure 5.

3.2. Description of the Neural Network Approach. The proposed NN is a multilayer network of (6-6-1) whose adopted architecture is illustrated in Figures 6 and 7. Note that the smallest error is obtained after 38 iterations.

Each neuron is connected to all the neurons of the next layer by connections whose weights are any real numbers.

The neuron network study used in this paper is carried out through the three main steps:

- (i) The construction of the network NN block using the Levenberg algorithm

- (ii) The data acquisition (learning base)
- (iii) The network test

3.3. Construction of the NN Block System. Figure 6 shows that our network consists of three layers:

- (i) An input layer composed of six neurons, whose role is to transmit the values of the inputs that correspond to the maxima of details ($\max(d1)$, $\max(d2)$, $\max(d3)$, $\max(d4)$, $\max(d5)$, and $\max(d6)$) to the next layer, called hidden layer
- (ii) A hidden layer with six neurons with selected sigmoid activation functions
- (iii) An output layer, which is composed of a neuron, where output of each neuron is 0 or 1

3.4. Acquisition of the Data (Learning Basis). Before building the NN block system, one must first access the learning phase. This can be in the form of a table. The latter consists of vectors (which represent the input layer of the NN), where each vector consists of 2 parameters.

A very rich database for healthy and faulty (open-circuit) cases can be developed, which has a lot of information about the open-circuit fault. During this phase, the maxima details of the healthy case are taken as the references; then the maxima of details for the faulty case are extracted and

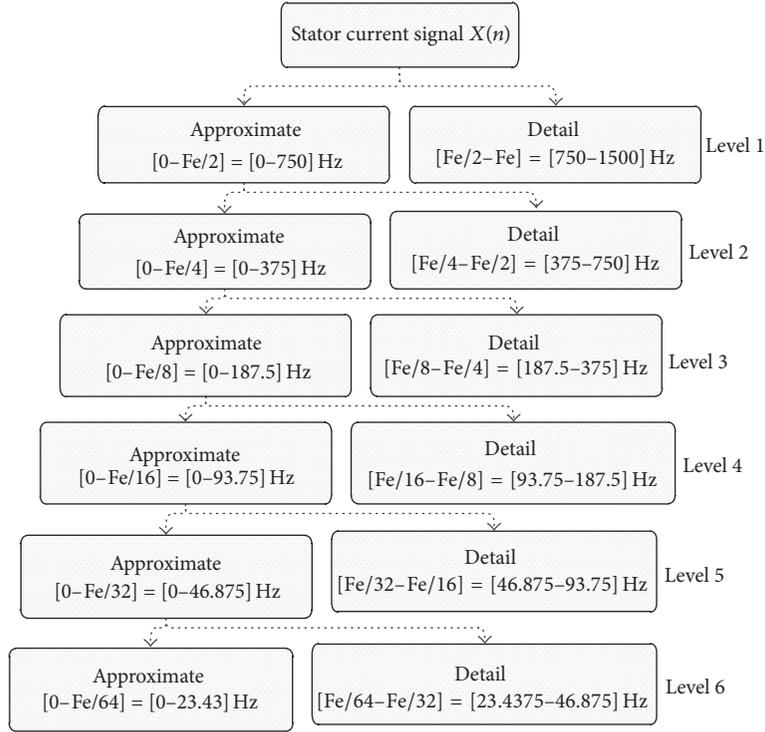


FIGURE 5: Decomposition of stator current signal $x(t)$.

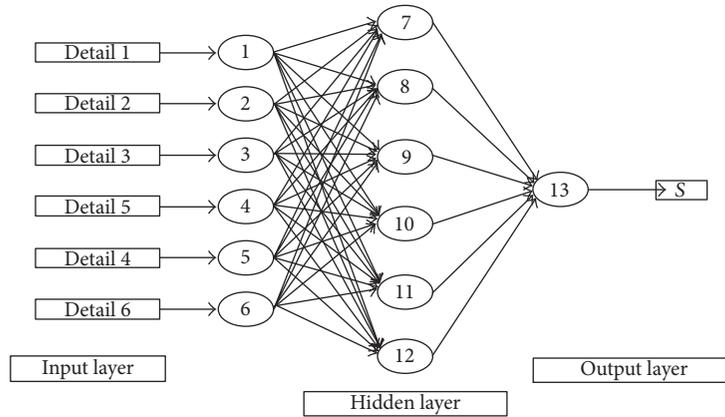


FIGURE 6: Proposed neural network architecture.

TABLE 1: Fault classification.

| State | Fault type | Symbol | Code |
|-------|-----------------------|--------|------|
| 1 | Healthy state | HS | 0 |
| 2 | Open-circuit at K_1 | OC | 1 |

compared with the healthy case. From this comparison, it can be deduced as either state 0 (i.e., no variation in detail) or state 1 (i.e., variation in detail). Table 1 is to resume the task of this phase.

3.5. *Network Test Results.* An automatic learning was performed using the MATLAB software. The learning is reached

once a small quadratic error of value $2.2181e^{-16}$ is obtained; see Figure 7. Note that the smallest error is obtained after 38 iterations.

4. Experimental Test-Rig

4.1. *Experimental Test-Rig Description.* The experimental test-rig used in this paper work includes a three-phase induction squirrel-cage motor fed by a three-phase two-level voltage source inverter. The detailed characteristics of the motor are given in the Appendix. Furthermore, the motor is mechanically coupled to a DC generator supplying resistors which allows varying the load torque. Moreover, the

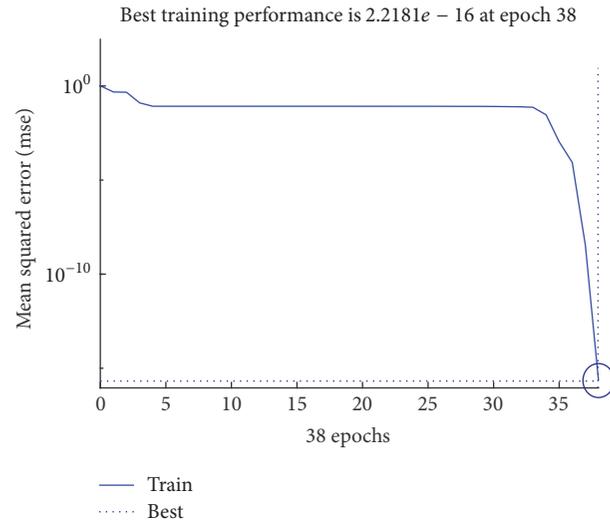


FIGURE 7: Evaluation of the quadratic error as a function of the number of learning iterations (using the method of retrogradient propagation).

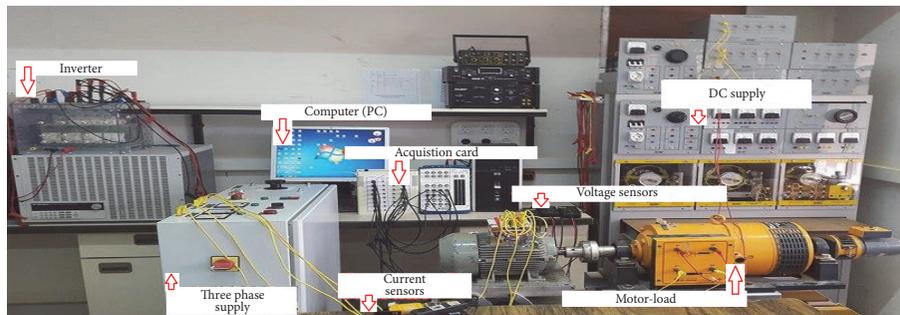


FIGURE 8: Photo of the experimental test-rig.

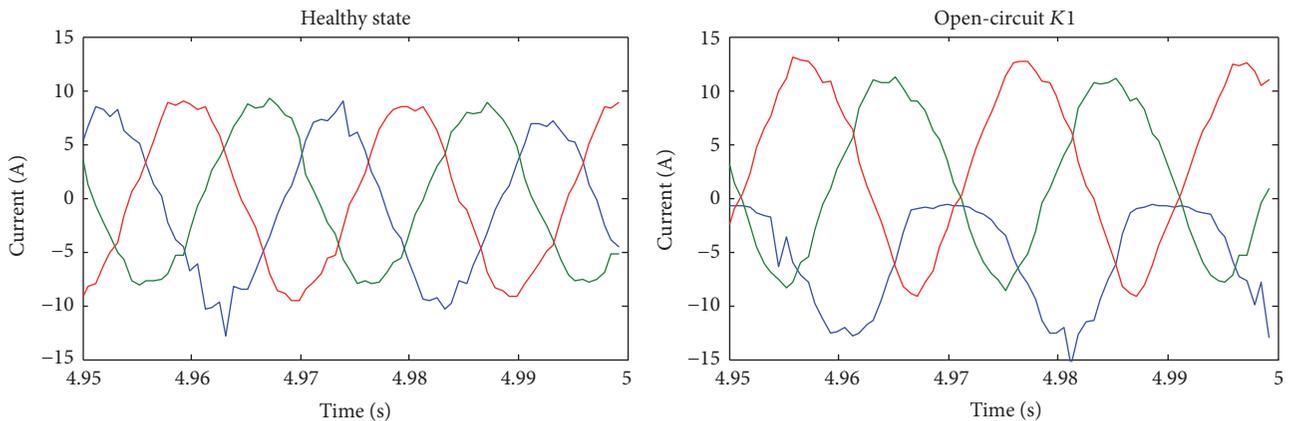


FIGURE 9: Currents waveforms of an induction motor for healthy and faulty open-circuit K_1 case.

measuring system includes three current Hall effect sensors and three voltage sensors and a DSPACE 1104 acquisition card to generate pulses for triggering the IGBTs gates. The whole set is connected to a computer for visualizing the processed sensed signal as shown in the photo of Figure 8. The acquisition time is taken as $T_{acq} = 5$ s and the sampling frequency $F_e = 1500$ Hz.

4.2. Experimental Results Presentation and Discussion. Figure 9 depicts the phase current waveforms of the induction motor for both a healthy state and an IGBT open-circuit faulty inverter.

From the experimental results depicting the currents waveforms of the motor in Figure 9, following an IGBT open-circuit fault K_1 of the inverter leg, the phase current

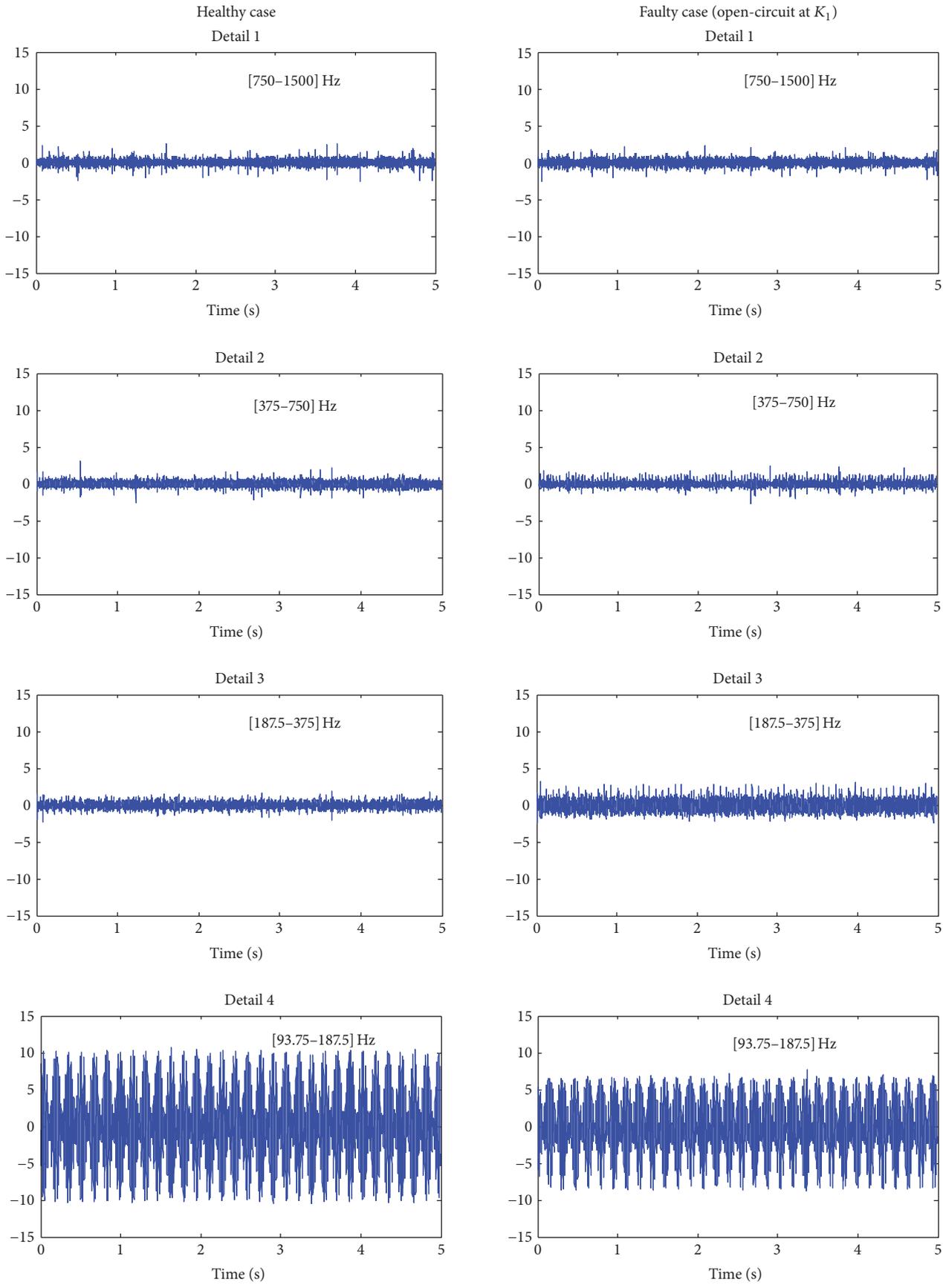


FIGURE 10: Continued.

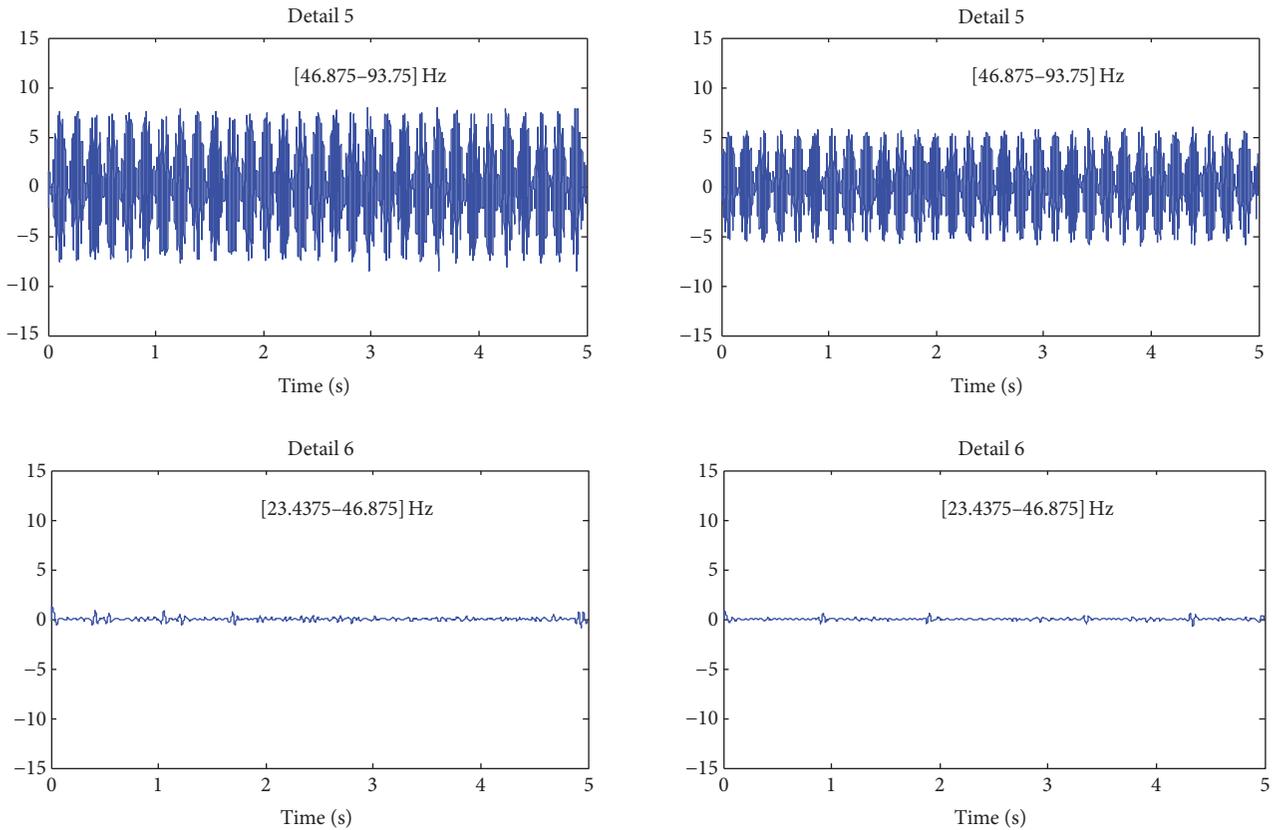


FIGURE 10: Details signal decomposition for healthy and faulty open-circuit at K_1 cases.

connected to this leg can no longer be controlled as it can only be negative or zero. The sum of the currents of the other two healthy phases is zero which may make it impossible to start the motor or to ensure for long term the continuity of service of the motor.

Figure 10 shows the various details extracted from the acquired stator current signal using the DWT technique for the healthy case and the faulty (open-circuit switch at K_1) case.

Figure 11 depicts the fault diagnostic system and Table 2 shows some examples of the diagnostic results.

Table 2 presents the input and output of the NN.

By comparing the details for the case of the healthy inverter state and that of the open-circuit switch fault state as depicted in Figure 11 or Table 2, a noticeable change in the amplitude of both details d_3 , d_4 is easily observed, while the rest of the details remain almost the same independently of the open-circuit switch fault presence or not. The change in these two details d_3 and d_4 indicates the existence of certain information within our stator current signal. To extract and explain this information, a stator current spectral analysis based on the FFT is performed for both details.

Figure 12 presents the spectra of the frequency ranges d_3 : [187.5-375] and d_4 : [93.75-187.5] for the healthy and the faulty cases, respectively.

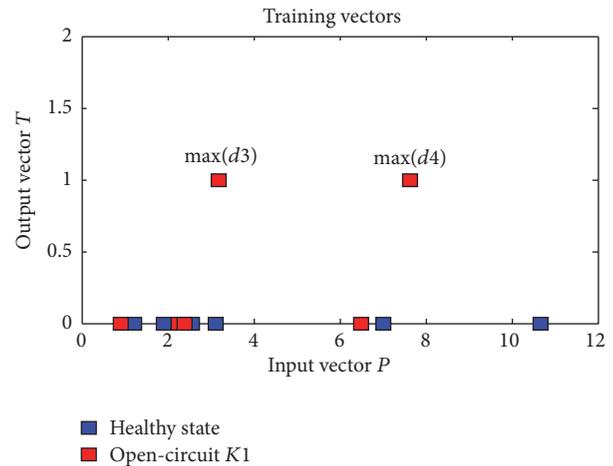


FIGURE 11: Variation of the details maxima.

Comparing both spectra in Figure 12 for healthy state and open-circuit fault state, we can easily notice the presence of additional harmonics characterizing the inverter open-circuit fault at frequencies 100 Hz in the first band [93.75-187.5] and 200 Hz and 300 Hz in the second band [187.5-375]. The amplitude and frequency of the various

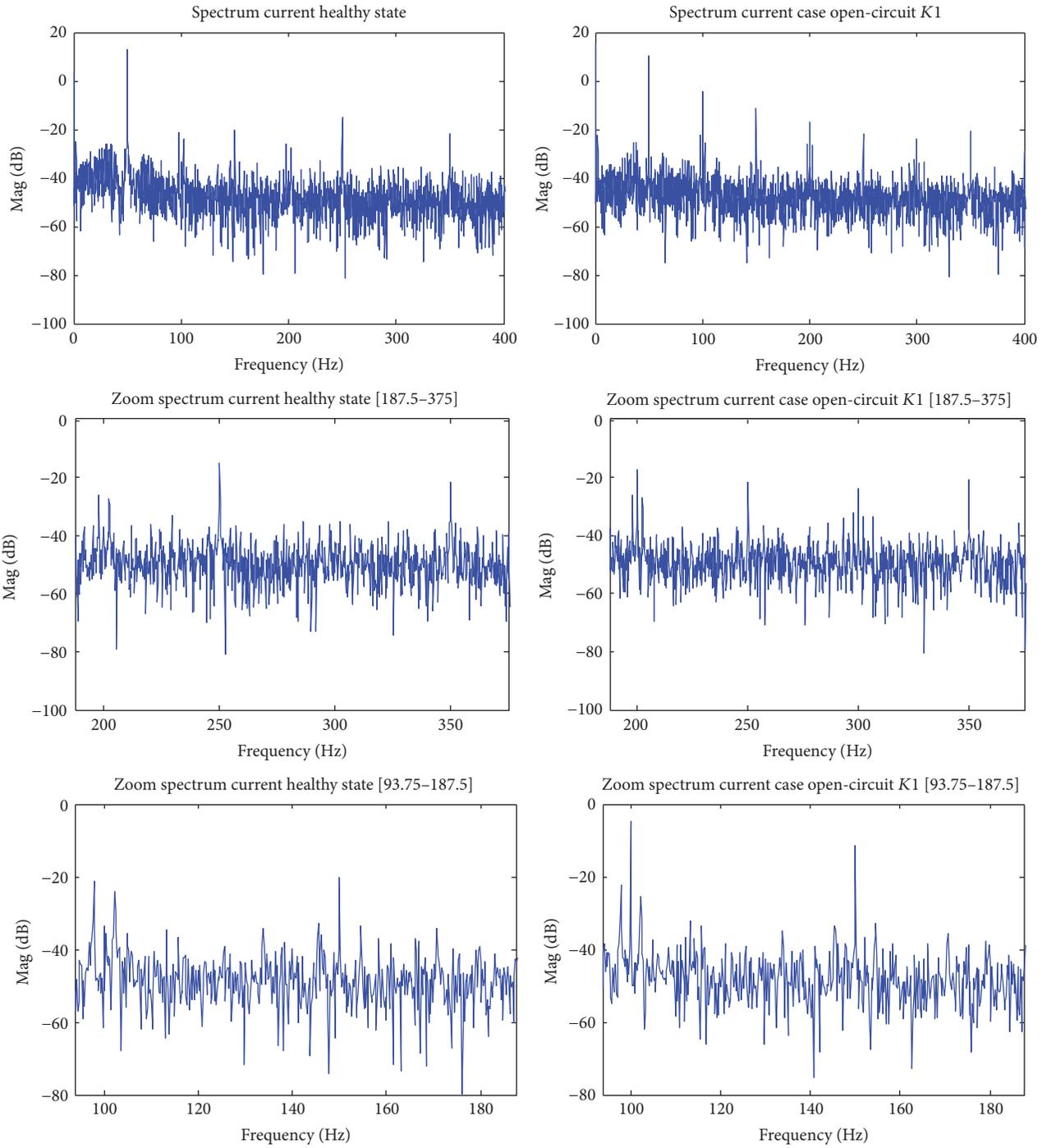


FIGURE 12: FFT of stator current signal for healthy and faulty open-circuit at K_1 cases.

TABLE 2: Details of one stator current phase for both healthy and faulty open-circuit at K_1 cases.

| Max details | $\max(d_1)$ | $\max(d_2)$ | $\max(d_3)$ | $\max(d_4)$ | $\max(d_5)$ | $\max(d_6)$ |
|--------------------------|-------------|-------------|-------------|-------------|-------------|-------------|
| Input | | | | | | |
| Healthy state | 2.568 | 3.126 | 1.928 | 10.67 | 7.014 | 1.228 |
| Open-circuit state K_1 | 2.261 | 2.412 | 3.206 | 7.632 | 6.502 | 0.909 |
| Output | -0.0000 | 0.0000 | 1.0000 | 1.0000 | 0.0000 | 0.0000 |

TABLE 3: Amplitudes and frequencies of harmonics.

| Harmonic (Hz) | $f_{oc} = 2f_s = 100$ Hz | $4f_s = 200$ Hz | $6f_s = 300$ Hz |
|-------------------------|--------------------------|-----------------|-----------------|
| Healthy state (db) | -33.51 db | -37.25 db | 0 db |
| Open-circuit state (db) | -4.623 db | -17.26 db | -23.77 db |

obtained harmonics for the healthy and faulty cases are summarized in Table 3.

A comparative analysis between both healthy and faulty states shows with more clarity a particular frequency signature around 100 Hz for the spectrum level $d3$. Note that the open-circuit frequency $f_{oc} = 2f_s = 100$ Hz corresponds to the harmonic frequency that characterizes the open-circuit fault of the IGBT switch.

5. Conclusion

In this paper, a research area dealing with the technique of diagnosis and detection of open-circuit fault in a three-phase two-level voltage source inverter fed induction motor is investigated. The paper proposes a diagnosis approach based on the association of both the discrete wavelet transform (DWT) and the neural network (NN) for the detection of the IGBT open-circuit fault of an inverter.

The study focuses first on the extraction of the details for the cases of the healthy and the open-circuit faulty IGBT by using the DWT algorithm. The investigation of the harmonics related to the obtained details particularly $d3$ and $d4$ is then conducted by using the FFT technique. The NN enables the development of a rich database for both healthy and faulty cases resulting in the automation of the diagnostic system.

The various obtained results are validated by several experimental tests conducted in the LDEE laboratory by the group diagnostic to assess the effectiveness and merits of the combined DWT-NN proposed approach.

Appendix

Rated power: 3 KW

Supply frequency: 50 Hz

Rated voltage: 380 V

Rated current: 7 A

Rotor speed: 1440 rev/min

Number of rotor bars: 28

Number of stator slots: 36

Power factor: 0.83

Number of pair of poles: 2

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

References

- [1] W. Zhang, D. Xu, P. N. Enjeti, H. Li, J. T. Hawke, and H. S. Krishnamoorthy, "Survey on fault-tolerant techniques for power electronic converters," *IEEE Transactions on Power Electronics*, vol. 29, no. 12, pp. 6319–6331, 2014.
- [2] T. Orłowska-Kowalska and P. Sobanski, "Simple diagnostic technique of a single IGBT open-circuit faults for a SVM-VSI vector controlled induction motor drive," *Bulletin of the Polish Academy of Sciences—Technical Sciences*, vol. 63, no. 1, pp. 281–288, 2015.
- [3] B. D. E. Cherif, A. Bendiabdellah, and M. A. Khelif, "Detection of open-circuit fault in a three-phase voltage inverter fed induction motor," *International Review of Automatic Control*, vol. 9, no. 6, pp. 374–382, 2016.
- [4] B. D. E. Cherif, M. Bendjebbar, N. Benouzza, H. Boudinar, and A. Bendiabdellah, "A comparative study between two open-circuit fault detection and localization techniques in a three-phase inverter fed induction motor," in *Proceedings of the 2016 8th International Conference on Modelling, Identification and Control (ICMIC)*, pp. 1–7, Algiers, November 2016.
- [5] H. Keskes, A. Braham, and Z. Lachiri, "Broken rotor bar diagnosis in induction machines through stationary wavelet packet transform and multiclass wavelet SVM," *Electric Power Systems Research*, vol. 97, pp. 151–157, 2013.
- [6] R. Yan, R. X. Gao, and X. Chen, "Wavelets for fault diagnosis of rotary machines: a review with applications," *Signal Processing*, vol. 96, pp. 1–15, 2013.
- [7] A. F. Aimer, A. H. Boudinar, N. Benouzza, and A. Bendiabdellah, "Simulation and experimental study of induction motor broken rotor bars fault diagnosis using stator current spectrogram," in *Proceedings of the 3rd International Conference on Control, Engineering and Information Technology, CEIT 2015*, dza, May 2015.
- [8] Z. Liu, Y. Cui, and W. Li, "Combined power quality disturbances recognition using wavelet packet entropies and S-transform," *Entropy*, vol. 17, no. 8, pp. 5811–5828, 2015.
- [9] Q. Yang and J. Wang, "Multi-level wavelet shannon entropy-based method for single-sensor fault location," *Entropy*, vol. 17, no. 10, pp. 7101–7117, 2015.
- [10] S. Kaur, K. Gaganpreet, and S. Dheerendra, "Comparative analysis of haar and coiflet wavelets using discrete wavelet transform in digital image compression," *International Journal of Computer Applications (IJERA)*, vol. 3, no. 3, pp. 669–673, 2013.
- [11] J. A. Antonino-Daviu, M. Riera-Guasp, J. R. Folch, and M. P. Molina Palomares, "Validation of a new method for the diagnosis of rotor bar failures via wavelet transform in industrial induction machines," *IEEE Transactions on Industry Applications*, vol. 42, no. 4, pp. 990–996, 2006.
- [12] R. R. Errabelli and P. Mutschler, "Fault-tolerant voltage source inverter for permanent magnet drives," *IEEE Transactions on Power Electronics*, vol. 27, no. 2, pp. 500–508, 2012.
- [13] T. Orłowska-Kowalska and P. Sobanski, "Simple sensorless diagnosis method for open-switch faults in SVM-VSI-fed induction motor drive," in *Proceedings of the 39th Annual Conference of the IEEE Industrial Electronics Society, IECON 2013*, pp. 8210–8215, Austria, November 2013.
- [14] W.-S. Im, J.-S. Kim, J.-M. Kim, D.-C. Lee, and K.-B. Lee, "Diagnosis methods for IGBT open switch fault applied to 3-phase AC/DC PWM converter," *Journal of Power Electronics*, vol. 12, no. 1, pp. 120–127, 2012.

- [15] I. Jlassi, S. Khojet, and E. Khil, "A MRAS-Luenberger observer based fault tolerant control of PMSM drive," *Journal of Electrical Systems*, vol. 10, no. 1, pp. 48–62, 2014.
- [16] B. Fan, Y. Yin, and C. Fu, "A method of inverter circuit fault diagnosis based on BP neural network and D-S evidence theory," in *Proceedings of the 2010 8th World Congress on Intelligent Control and Automation, WCICA 2010*, pp. 2249–2253, China, July 2010.
- [17] L. hong and W. yan-qiu, "Study of Fault Diagnosis on Three-phase Sine-PWM Inverter Based on Rough Set-neural Network System," *Journal of Liaoning Istitute of Technology*, vol. 25, no. 6, pp. 351–353, 2005.

Research Article

Neural Back-Stepping Control of Hypersonic Flight Vehicle with Actuator Fault

Qi Wu ¹ and Yuyan Guo ²

¹Department of Computer Science, Guangdong Police College, Guangzhou 510230, China

²School of Automation, Northwestern Polytechnical University, Xi'an 710072, China

Correspondence should be addressed to Qi Wu; wuqi888@126.com

Received 14 October 2017; Accepted 3 January 2018; Published 1 March 2018

Academic Editor: Youqing Wang

Copyright © 2018 Qi Wu and Yuyan Guo. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This paper addresses the fault-tolerant control of hypersonic flight vehicle. To estimate the unknown function in flight dynamics, neural networks are employed in controller design. Moreover, in order to compensate the actuator fault, an adaptive signal is introduced in the controller design to estimate the unknown fault parameters. Simulation results demonstrate that the proposed approach could obtain satisfying performance.

1. Introduction

Fault diagnosis and fault-tolerant control have been a hot issue since its significance on actual systems with actuator faults. By identifying or estimating unknown faults, the influence of faults could be considered and eliminated in controller design. Representative study on this issue could be found in [1, 2], while in this paper, fault diagnosis have been studied and applied on hypersonic flight vehicle (HFV) whose control methods have attracted a lot of attention. Hypersonic flight vehicle has a much higher speed than traditional aircraft, however due to its special flight environment as well as structure, the controller design of HFV faces more challenges [3].

Currently, most studies on HFV control did not take the actuator fault into consideration. However, unknown fault such as actuator dead-zone is quite common in nonlinear systems and may cause serious consequences [4]. In order to remove the influence caused by actuator fault, several approaches have been studied such as fuzzy control [5], robust control [6, 7], and adaptive compensation [8, 9]. The model of the HFV also contains uncertain nonlinearity. As a result, studies on HFV have concentrated on the estimation of the uncertainty. Among all the methods, neural

network has been widely used due to its good performance in approximating unknown nonlinear functions [10].

The model of HFV is highly nonlinear; meanwhile the altitude subsystem could be considered as a strict-feedback system [11]. Therefore, back-stepping method [12–14] in HFV control has been widely studied. To eliminate continuous derivatives of each virtual control in back-stepping design, multiple methods such as dynamic surface control [15] and differentiator [14] are combined with back-stepping which makes the approach practicable. It is noticed that in [16] the multiple actuator fault is considered and the robust design is presented to guarantee the system stability. Moreover, in [17], the input dead-zone is considered where the Nussbaum design is included to make the adaptive design available. In this paper, we try to estimate the parameters of system fault in the hypersonic flight dynamics and then the “dynamic inversion” of the actuator fault can be included in the control signal. The whole design is using a back-stepping control law with neural networks and adaptive method.

The paper is organized in 6 parts. In Section 2, the control-oriented model (COM) of HFV considered in this paper is simply introduced. In Sections 3 and 4, the back-stepping controller based on neural networks and adaptive fault estimation is designed and system stability is analysed.

The simulation results of the designed controller are shown in Section 5. Finally the summary is given in Section 6.

2. Longitudinal Dynamics of HFV with Actuator Fault

The COM in [18] is employed for study:

$$\begin{aligned}\dot{V} &= \frac{T \cos \alpha - D}{m} - g \sin \gamma \\ \dot{h} &= V \sin \gamma \\ \dot{\gamma} &= \frac{L + T \sin \alpha}{mV} - \frac{g \cos \gamma}{V} \\ \dot{\alpha} &= q - \dot{\gamma} \\ \dot{q} &= \frac{M_{yy}}{I_{yy}}.\end{aligned}\quad (1)$$

The detail of the dynamics can be found in [18]. Consider the following actuator fault model:

$$\delta_e = \begin{cases} u - b & \text{if } u \geq b \\ 0 & \text{if } -b < u < b \\ u + b & \text{if } u \leq -b, \end{cases}\quad (2)$$

where u is the designed control input and $b < 0$ is an unknown fault parameter to be estimated.

Remark 1. In HFV systems, the fault will cause error between designed and actual control input and it will result in tracking error or even flight instability.

3. Adaptive Back-Stepping Controller

The strict-feedback form altitude subsystem based on the hypersonic flight dynamics is considered

$$\begin{aligned}\dot{x}_1 &= f_1 + g_1 x_2 \\ \dot{x}_2 &= x_3 \\ \dot{x}_3 &= f_3 + g_3 \delta_e,\end{aligned}\quad (3)$$

where $x_1 = \gamma$, $x_2 = \gamma + \alpha$, and $x_3 = q$. f_i, g_i are nonlinear functions of the HFV model.

Assumption 2. In altitude subsystem (3), g_3 is a bounded function.

Flight path angle (FPA) tracking error \tilde{x}_1 is defined as

$$\tilde{x}_1 = x_1 - x_{1d},\quad (4)$$

where x_{1d} is the command signal of FPA which is designed via altitude reference signal.

Step 1. Choose virtual control of x_2 as

$$x_{2d} = \frac{1}{g_1} (-f_1 - c_1 \tilde{x}_1 + \dot{x}_{1d}),\quad (5)$$

where c_1 is the control gain.

To avoid the continuous derivative of virtual control, the following first-order differentiator is designed:

$$\begin{aligned}\dot{\hat{\omega}}_{2c} &= -l_{20} \sqrt{|\hat{\omega}_{2c} - x_{2d}|} \text{sgn}(\hat{\omega}_{2c} - x_{2d}) + \hat{\omega}_{2d} \\ \dot{\hat{\omega}}_{2d} &= -l_{21} \text{sgn}(\hat{\omega}_{2d} - \dot{\hat{\omega}}_{2c}),\end{aligned}\quad (6)$$

where l_{20} and l_{21} are positive designed parameters.

Step 2. Define the tracking error \tilde{x}_2 :

$$\tilde{x}_2 = x_2 - x_{2d}.\quad (7)$$

Choose virtual control of x_3 as

$$x_{3d} = -c_2 \tilde{x}_2 + \dot{\hat{\omega}}_{2c} - g_1 \tilde{x}_1.\quad (8)$$

where c_2 is the control gain.

The following first-order differentiator is designed:

$$\begin{aligned}\dot{\hat{\omega}}_{3c} &= -l_{30} \sqrt{|\hat{\omega}_{3c} - x_{3d}|} \text{sgn}(\hat{\omega}_{3c} - x_{3d}) + \hat{\omega}_{3d} \\ \dot{\hat{\omega}}_{3d} &= -l_{31} \text{sgn}(\hat{\omega}_{3d} - \dot{\hat{\omega}}_{3c}),\end{aligned}\quad (9)$$

where l_{30} and l_{31} are positive designed parameters.

Step 3. Define the tracking error \tilde{x}_3 :

$$\tilde{x}_3 = x_3 - x_{3d}.\quad (10)$$

Due to the uncertainty caused by imprecise model, the nonlinear function f_3 may be unknown. Therefore its NN-based estimation value is employed in the control law:

$$u_0 = \frac{1}{g_3} (-\hat{\omega}_3^T \Theta_3 - c_3 \tilde{x}_3 - \tilde{x}_2 + \dot{\hat{\omega}}_{3c}),\quad (11)$$

where c_3 is the control gain and $\hat{\omega}_3$ is the estimation of the optimal NN weights ω_3^* , which is obtained via adaptive law:

$$\dot{\hat{\omega}}_3 = \Gamma_3 \tilde{x}_3 \Theta_3(\tilde{x}_3) - \Gamma_3 \delta_3 \hat{\omega}_3,\quad (12)$$

where Γ_3 and δ_3 are positive parameters. Θ_3 is obtained via radial basis function. Define $\tilde{\omega}_3 = \omega_3^* - \hat{\omega}_3$.

In order to eliminate the influence of unknown constant b , signal \hat{b} is introduced in controller design to compensate the actuator:

$$u = u_0 + \hat{b} \text{sgn}(u_0).\quad (13)$$

The adaptive law is designed as

$$\dot{\hat{b}} = -\Gamma_b g_3 \tilde{x}_3 \text{sgn}(u_0) - \Gamma_b \sigma_b \hat{b},\quad (14)$$

where Γ_b and σ_b are positive designed parameters. Define the estimation error $\tilde{b} = b - \hat{b}$.

Remark 3. In previous work on HFV dead-zone fault control [17], the paper regarded the dead-zone as a part of the compound disturbance, where robust technique is employed. In our paper, we proposed an adaptive law to estimate the unknown fault parameter and added a compensating signal in the controller so that the influence of dead-zone fault could be directly eliminated, which is shown by Figures 3 and 5 in our manuscript.

Define the velocity tracking error as

$$\tilde{V} = V - V_r, \quad (15)$$

where V_r is the reference signal. Then the following PID controller is designed:

$$\Phi = k_{pv}\tilde{V} + k_{iv} \int \tilde{V}(t) dt + k_{dv}\dot{\tilde{V}}. \quad (16)$$

4. Stability Analysis

Theorem 4. Consider the HFV altitude system (3) with control laws (5), (8), (11), and (13) and adaptive laws (12) and (14); all of the error signals are uniformly ultimately bounded.

Proof. The Lyapunov function candidate is chosen as

$$V_L = \sum_{i=1}^3 V_i, \quad (17)$$

where

$$\begin{aligned} V_1 &= \frac{1}{2}\tilde{x}_1^2 \\ V_2 &= \frac{1}{2}\tilde{x}_2^2 \\ V_3 &= \frac{1}{2}\tilde{x}_3^2 + \frac{1}{2}\tilde{\omega}_3^T \Gamma_3^{-1} \tilde{\omega}_3 + \frac{1}{2}\Gamma_b^{-1} \tilde{b}^2. \end{aligned} \quad (18)$$

The derivative of V_L is obtained as

$$\dot{V}_L = \sum_{i=1}^3 \tilde{x}_i (\dot{x}_i - \dot{x}_{id}) - \tilde{\omega}_3^T \Gamma_3^{-1} \dot{\tilde{\omega}}_3 - \Gamma_b^{-1} \dot{\tilde{b}}. \quad (19)$$

According to the conclusion in [19], $\hat{\omega}_{2c}$, $\hat{\omega}_{3c}$ in differentiators (6), (9) could estimate \dot{x}_{2d} and \dot{x}_{3d} to arbitrary accuracy. Therefore there exists

$$\dot{x}_{id} = \hat{\omega}_{ic} + \chi_i, \quad i = 2, 3, \quad (20)$$

where $\chi_i \leq |\chi_{im}|$, χ_{im} is a positive constant. Substitute the control laws and adaptive laws into (19); there exists

$$\begin{aligned} \dot{V}_L &= -\sum_{i=1}^3 \varsigma_i \tilde{x}_i^2 + \delta_3 (\tilde{\omega}_3^T \omega_3^* - \tilde{\omega}_3^T \tilde{\omega}_3) + \varepsilon_3 \tilde{x}_3 \\ &+ b \tilde{x}_3 g_3 \left[\text{sgn}(u_0) - \text{sgn}(u) \right] + \sigma_b (b\tilde{b} - \tilde{b}^2) \\ &- \sum_{i=2}^3 \tilde{x}_i \chi_i, \end{aligned} \quad (21)$$

where ε_3 is the bounded inevitable NN reconstruction error satisfying $|\varepsilon_3| \leq \varepsilon_{3m}$, where ε_{3m} is a positive constant. Define $\eta = bg_3[\text{sgn}(u_0) - \text{sgn}(u)]$; consider Assumption 1; η is a bounded signal satisfying $|\eta| \leq \eta_m$, where η_m is a positive constant.

The following inequalities hold:

$$\begin{aligned} \delta_3 (\tilde{\omega}_3^T \omega_3^* - \tilde{\omega}_3^T \tilde{\omega}_3) &\leq \frac{1}{2} \delta_3 (\|\omega_3^*\|^2 - \tilde{\omega}_3^T \tilde{\omega}_3) \\ \sigma_b (b\tilde{b} - \tilde{b}^2) &\leq \frac{1}{2} \sigma_b (b^2 - \tilde{b}^2) \\ -\sum_{i=2}^3 \tilde{x}_i \chi_i &\leq \sum_{i=2}^3 \left(\frac{1}{2} \tilde{x}_i^2 + \frac{1}{2} \chi_{im}^2 \right) \\ \varepsilon_3 \tilde{x}_3 &\leq \frac{1}{2} \varepsilon_{3m}^2 + \frac{1}{2} \tilde{x}_3^2 \\ \eta \tilde{x}_3 &\leq \frac{1}{2} \eta_m^2 + \frac{1}{2} \tilde{x}_3^2. \end{aligned} \quad (22)$$

The derivative of V_L satisfies

$$\begin{aligned} \dot{V}_L &\leq -\varsigma_1 \tilde{x}_1^2 - \left(\varsigma_2 - \frac{1}{2} \right) \tilde{x}_2^2 - \left(\varsigma_3 - \frac{3}{2} \right) \tilde{x}_3^2 - \frac{1}{2} \delta_3 \tilde{\omega}_3^T \tilde{\omega}_3 \\ &- \frac{1}{2} \sigma_b \tilde{b}^2 + \frac{1}{2} \delta_3 \|\omega_3^*\|^2 + \frac{1}{2} \varepsilon_{3m}^2 + \frac{1}{2} \eta_m^2 \\ &+ \frac{1}{2} \sigma_b b^2 + \sum_{i=2}^3 \frac{1}{2} \chi_{im}^2 \\ &\leq -\sigma V_L + P, \end{aligned} \quad (23)$$

where $\sigma = \min\{2\varsigma_1, (2\varsigma_2 - 1), (2\varsigma_2 - 3), \delta_3 \Gamma_3, \sigma_b \Gamma_b\}$ and $P = (1/2)\delta_3 \|\omega_3^*\|^2 + (1/2)\varepsilon_{3m}^2 + (1/2)\eta_m^2 + (1/2)\sigma_b b^2 + \sum_{i=2}^3 (1/2)\chi_{im}^2$. Then all of the signals in V_L are uniformly ultimately bounded. This completes the proof. \square

5. Simulation

Let altitude increase 152.4 m from the initial value; notice that the signal will pass the following filter to make sure the reference signal is smooth enough for tracking:

$$\frac{\Delta_r}{\Delta} = \frac{\omega_1 \omega_2^2}{(s + \omega_1)(s^2 + 2\xi \omega_2 s + \omega_2^2)}, \quad (24)$$

where $\omega_1 = 0.8$, $\omega_2 = 0.5$, and $\xi = 0.1$. The FPA command signal is obtained as

$$x_{1d} = \arcsin \left(\frac{-k_h \tilde{h} - k_i \int \tilde{h} dt + \dot{h}_r}{V} \right), \quad (25)$$

where $k_h = 0.5$, $k_i = 0.05$, $\tilde{h} = h - h_r$ is the altitude tracking error, and h_r is the altitude reference signal.

The parameters in the controller are set as $\varsigma_1 = 2$, $\varsigma_2 = 2$, $\varsigma_3 = 3$, $k_{pv} = 5$, $k_{dv} = 0.01$, and $k_{iv} = 0.01$. The parameters of first-order differentiator and adaptive laws are chosen as $l_{20} = 120$, $l_{30} = 100$, $l_{21} = 0.9$, $l_{31} = 0.05$, $\Gamma_3 = 0.5$, $\delta_3 = 0.002$,

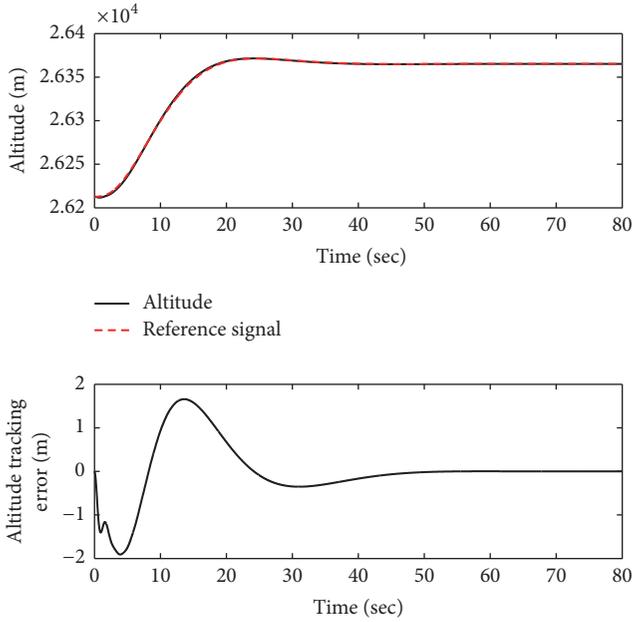


FIGURE 1: Altitude.

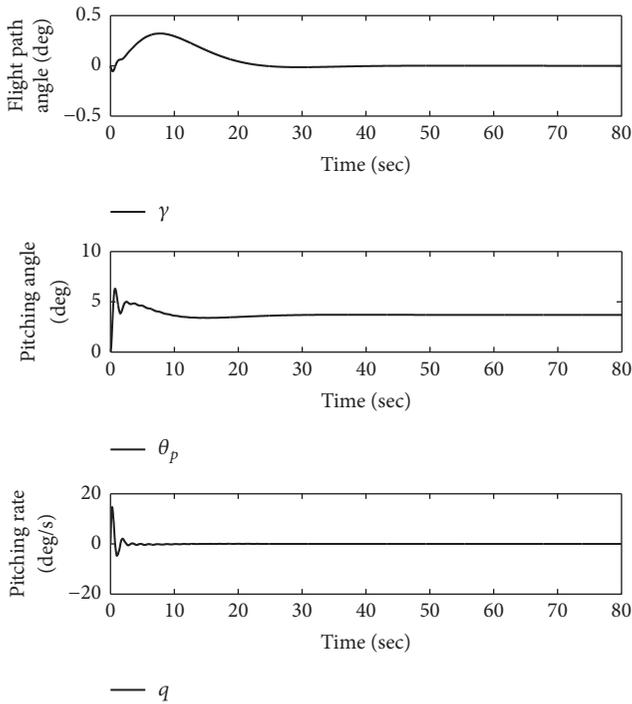


FIGURE 2: FPA, pitching angle, and pitching rate.

$\Gamma_b = 0.113$, and $\sigma_b = 0.005$. Take the initial states as $h_0 = 26212$ m, $v_0 = 2743$ m/s, $\gamma_0 = 0$ deg, $\alpha_0 = 0$ deg, and $q_0 = 0$ deg/s. The fault parameter is set as $b = 0.05$. The results are as follows.

Figure 1 indicates that the designed controller could obtain good performance on trajectory tracking. Figure 2 shows that all state variables are bounded. The response of the designed adaptive signal \hat{b} and NN weights norm are shown

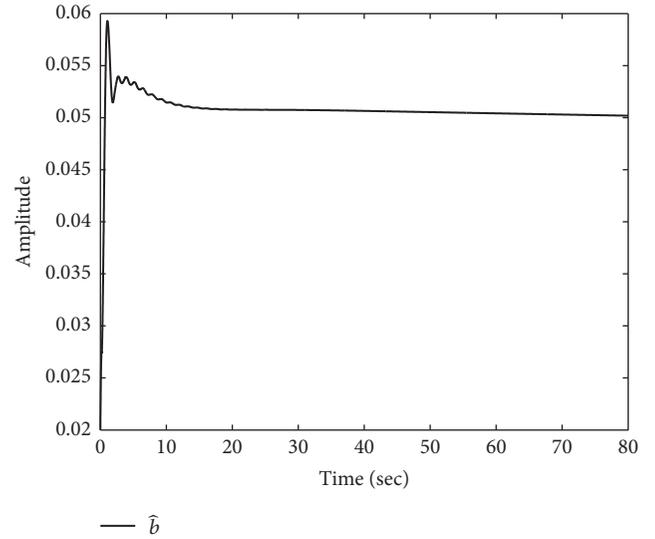
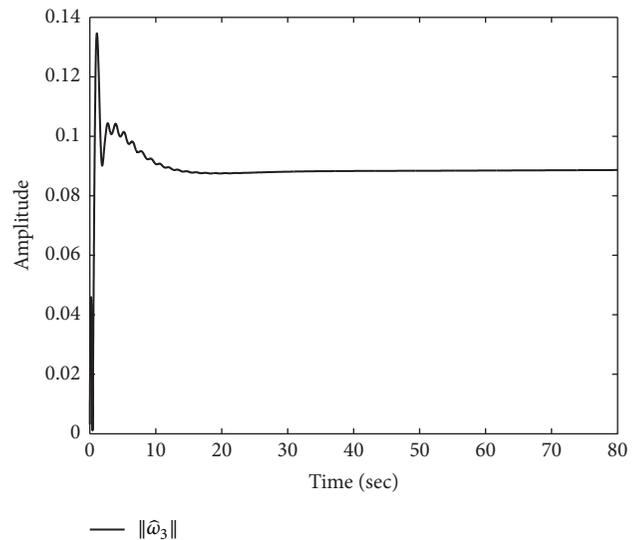
FIGURE 3: Response of \hat{b} .

FIGURE 4: NN weights norm.

in Figures 3 and 4. The designed control input and actual control input are shown in Figure 5, respectively. The results verify that the adaptive compensation design could eliminate the influence of actuator fault.

6. Conclusion

This paper proposes an adaptive back-stepping control law with NN learning for HFV control. The influence of actuator fault is eliminated by constructing an adaptive compensation signal. Meanwhile, the unknown nonlinearity is estimated by neural networks. The simulation results clearly present the consequence of the above design and verify that the approach could reach the desired tracking performance when actuator fault and model uncertainty exist.

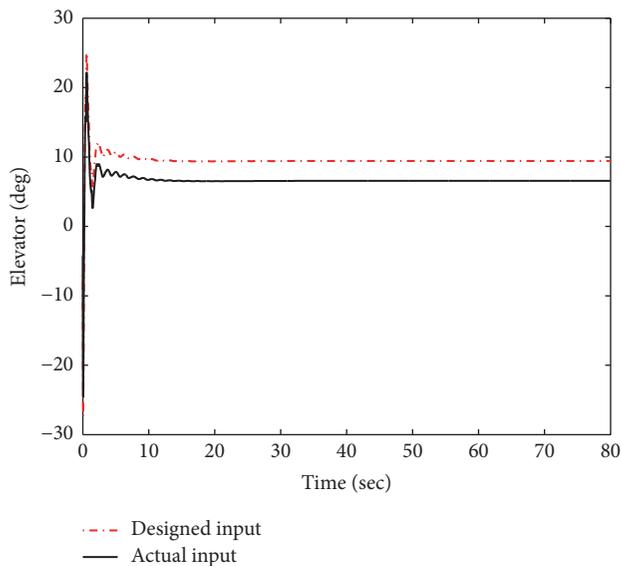


FIGURE 5: Control input.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

References

- [1] P. M. Frank, "Fault diagnosis in dynamic systems using analytical and knowledge-based redundancy: a survey and some new results," *Automatica*, vol. 26, no. 3, pp. 459–474, 1990.
- [2] D. Xu, B. Jiang, and P. Shi, "Robust NSV fault-tolerant control system design against actuator faults and control surface damage under actuator dynamics," *IEEE Transactions on Industrial Electronics*, vol. 62, no. 9, pp. 5919–5928, 2015.
- [3] H. Xu, M. D. Mirmirani, and P. A. Ioannou, "Adaptive sliding mode control design for a hypersonic flight vehicle," *Journal of Guidance, Control, and Dynamics*, vol. 27, no. 5, pp. 829–838, 2004.
- [4] S. Ibrir, W. F. Xie, and C.-Y. Su, "Adaptive tracking of nonlinear systems with non-symmetric dead-zone input," *Automatica*, vol. 43, no. 3, pp. 522–530, 2007.
- [5] Q. Shen, B. Jiang, and V. Cocquempot, "Fault diagnosis and estimation for near-space hypersonic vehicle with sensor faults," *Proceedings of the Institution of Mechanical Engineers, Part I: Journal of Systems and Control Engineering*, vol. 226, no. 3, pp. 302–313, 2012.
- [6] Y. Li, S. Tong, Y. Liu, and T. Li, "Adaptive fuzzy robust output feedback control of nonlinear systems with unknown dead zones based on a small-gain approach," *IEEE Transactions on Fuzzy Systems*, vol. 22, no. 1, pp. 164–176, 2014.
- [7] D. Zhao, Y. Wang, Y. Li, and S. X. Ding, " H_{∞} fault estimation for 2-D linear discrete time-varying systems based on krein space method," *IEEE Transactions on Systems, Man, and Cybernetics Systems*, vol. PP, no. 99, Article ID 2723623, pp. 1–10, 2017.
- [8] D. Zhao, D. Shen, and Y. Wang, "Fault diagnosis and compensation for two-dimensional discrete time systems with sensor faults and time-varying delays," *International Journal of Robust and Nonlinear Control*, vol. 27, no. 16, pp. 3296–3320, 2017.
- [9] Y. Wang, D. Zhao, Y. Li, and S. . Ding, "Unbiased minimum variance fault and state estimation for linear discrete time-varying two-dimensional systems," *Institute of Electrical and Electronics Engineers Transactions on Automatic Control*, vol. 62, no. 10, pp. 5463–5469, 2017.
- [10] S. Zhang, C. Li, and J. Zhu, "Composite dynamic surface control of hypersonic flight dynamics using neural networks," *Science China Information Sciences*, vol. 58, no. 7, 2015.
- [11] J.-L. Wu, "Stabilizing controllers design for switched nonlinear systems in strict-feedback form," *Automatica*, vol. 45, no. 4, pp. 1092–1096, 2009.
- [12] B. Xu, X. Huang, D. Wang, and F. Sun, "Dynamic surface control of constrained hypersonic flight models with parameter estimation and actuator compensation," *Asian Journal of Control*, vol. 16, no. 1, pp. 162–174, 2014.
- [13] B. Xu, D. Wang, Y. Zhang, and Z. Shi, "DOB based neural control of flexible hypersonic flight vehicle considering wind effects," *IEEE Transactions on Industrial Electronics*, vol. 64, no. 11, pp. 8676–8685, 2017.
- [14] B. Xu, D. Yang, Z. Shi, Y. Pan, B. Chen, and F. Sun, "Online recorded data-based composite neural control of strict-feedback systems with application to hypersonic flight dynamics," *IEEE Transactions on Neural Networks and Learning Systems*, vol. PP, no. 99, Article ID 2743784, pp. 1–11, 2017.
- [15] D. Swaroop, J. K. Hedrick, P. P. Yip, and J. . Gerdes, "Dynamic surface control for a class of nonlinear systems," *Institute of Electrical and Electronics Engineers Transactions on Automatic Control*, vol. 45, no. 10, pp. 1893–1899, 2000.
- [16] B. Xu, Y. Guo, Y. Yuan, Y. Fan, and D. Wang, "Fault-tolerant control using command-filtered adaptive back-stepping technique: application to hypersonic longitudinal flight dynamics," *International Journal of Adaptive Control and Signal Processing*, vol. 30, no. 4, pp. 553–577, 2016.
- [17] B. Xu, "Robust adaptive neural control of flexible hypersonic flight vehicle with dead-zone input nonlinearity," *Nonlinear Dynamics*, vol. 80, no. 3, pp. 1509–1520, 2015.
- [18] J. T. Parker, A. Serrani, S. Yurkovich, M. A. Bolender, and D. B. Doman, "Control-oriented modeling of an air-breathing hypersonic vehicle," *Journal of Guidance, Control, and Dynamics*, vol. 30, no. 3, pp. 856–869, 2007.
- [19] A. Levant, "Robust Exact Differentiation via Sliding Mode Technique," *Automatica*, vol. 34, no. 3, pp. 379–384, 1998.