IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received March 8, 2019, accepted March 17, 2019, date of publication March 25, 2019, date of current version April 5, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2906976

A Semi-Quantitative Information Based Fault Diagnosis Method for the Running Gears System of High-Speed Trains

CHAO CHENG[1,2,3], XINYU QIAO[1], HAO LUO[4], WANXIU TENG[2], MINGLIANG GAO[2,3], BANGCHENG ZHANG[5], AND XIAOJING YIN[5]

[1]School of Computer Science and Engineering, Changchun University of Technology, Changchun 130012, China
[2]National Engineering Laboratory, CRRC Changchun Railway Vehicles Co., Ltd., Changchun 130062, China
[3]Department of Automation, Tsinghua University, Beijing 100084, China
[4]Academy of Astronautics, Harbin Institute of Technology, Harbin 150001, China
[5]School of Mechatronic Engineering, Changchun University of Technology, Changchun 130012, China

Corresponding author: Bangcheng Zhang (zhangbangcheng@ccut.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 61751304 and Grant 61803044, in part by the Jilin Scientific and Technological Department Program under Grant 20180201125GX and Grant 20160520020JH, and in part by the Jilin Province Development and Reform Commission under Grant 2019C040-3.

ABSTRACT The proper operation of running gears of a high-speed train is one of the key factors to ensure its safety and reliability. The diagnosis of the state of running gears of a high-speed train is one of the effective ways to improve its reliability. It is difficult to diagnose the running gears of a high-speed train accurately because of the characteristics of its complex-analytic structure, multiple types of monitoring feature data, and lack of effective failure mode data. Therefore, this paper proposes a fault diagnosis method for the running gears of a high-speed train based on a semi-quantitative information model. The relation between the effective data and expert knowledge is studied, and the state of the running gears of a high-speed train is rigorously analyzed. To reduce the data dimension and the diagnostic calculation time of the running gears of a high-speed train, the principal component analysis (PCA) is used to screen its key monitoring features. Then, based on the change of the feature quantity in the working process of the running gears of a high-speed train, the semi-quantitative information model of belief-rule-base (BRB) fault diagnosis is established. In the diagnosis process, the initial model parameters of BRB are determined by expert knowledge and they have certain subjectivity. To improve the accuracy of the model, the constrained covariance matrix adaptive evolutionary strategy (CMA-ES) algorithm is used to optimize the parameters of the initial BRB model to improve the validity and accuracy of the diagnosis. Finally, to verify the effectiveness of the proposed semi-quantitative information model, a set of real data of the running gears of a high-speed train is used as case studies.

INDEX TERMS Semi-quantitative information, fault diagnosis, principle component analysis, belief-rule-base, constraint covariance matrix adaptive evolution strategy.

I. INTRODUCTION

Running gears are key parts of high-speed trains, and its operating condition directly affects the driving safety of a high-speed train [1]. The structure of a running gear of a high-speed train has a high degree of complexity, which includes traction drive, axle, axle box, a spring device, and a detection sensor. Any components of the high-speed train that appears to have pitting, indentation, exfoliation, or any other local defects can obviously cause it to sway on the carriage during its running process, and even cause rushing out of the track, which can result in unpredictable consequences. Therefore, it is crucial to improve the safety of the running gears of a high-speed train [2]–[4]. This paper studies the fault diagnosis of running gears to ensure that its operation is safe and reliable.

The running gears of a high-speed train are classified as a complex electromechanical system. The methods of fault diagnosis of complex electromechanical systems can be divided as follows: 1) qualitative analysis methods [5]–[7]: fault tree and expert system, 2) Data-driven methods [8]–[16]:

The associate editor coordinating the review of this manuscript and approving it for publication was Francesco Tedesco.

least squares method, support vector machine (SVM), neural network, and hidden Markov Method (HMM), 3) analytical model-based methods [17]–[20]: fuzzy reasoning. Swetapadma et al. [21] discussed a fault detection with a classification scheme based on the decision tree, and examined the faults in the faultless and the faulty state. Due to the subjective correlation coefficient of the method, the objective basis for the lack of data in the discriminant is considered. Miao et al. [22] studied a technique for fault diagnosis based on SVM and condition monitoring to detect the degree of degradation of the system. However, the method needs more accurate features, which affects the generalization ability of the system. Jiang et al. [23] analyzed the effect of variable selection on the monitoring performance of principal component analysis (PCA). They proposed a fault-relevant variable selection and Bayesian inference-based distributed method for efficient fault detection and isolation, which can provide an accurate description of faults. Huang et al. [24] proposed an improved HMM algorithm for urban rail transit motor equipment fault diagnosis. Its shortcoming was that the current state was only related to the previous state during the state transition, which could make the evaluation inaccurate. Zhao et al. [25] a new fault feature extraction method, called the EDOMFE method based on integrating ensemble empirical mode decomposition (EEMD), mode selection, and multiscale fuzzy entropy is proposed to accurately diagnose fault. In the above references, It is hard to establish an accurate analytical model for the running gears of a high-speed train due to the complexity of its structure. Additionally, it is difficult to accurately diagnose the fault because of multiple types of monitoring feature data and the lack of effective failure mode data.

To solve this problem, a fault diagnosis method for the running gears of a high-speed train based on semi-quantitative information model is proposed, which combines expert knowledge and monitoring data to establish the PCA−BRB model for the running gears of a high-speed train. The proposed BRB model is based on D-S evidence reasoning, IF-THEN rule expert system, and evidence reasoning method [7], [26]–[29]. It has a good ability to model nonlinear data with fuzzy uncertainty or probability uncertainty. The failure mode data of the running gears of a high-speed train provides a better solution to the fuzzy uncertainty or probability uncertainty. The PCA is the most representative dimension reduction model that screens the key premise attributes in the BRB model. It is necessary to simplify and improve the efficiency of the BRB model, which can effectively avoid the explosion of the BRB model combination. The existence of expert knowledge in the establishment of the initial BRB model is limited, and the model given by the experts is incomplete. To ensure the accuracy of the model and to establish an effective fault diagnosis model for the running gears of a high-speed train, the parameters are updated by applying the CMA-ES algorithm in the initial BRB model.

The rest of this paper is organized as follows. Section 2 introduces background studies of the problem and the motivation for applying semi-quantitative information model. The BRB algorithm and the optimization algorithm are discussed in section 3. In section 4, the running gears of the high-speed trains are discussed as an example and analysis of the proposed method. Section 5 presents the conclusion of this paper.

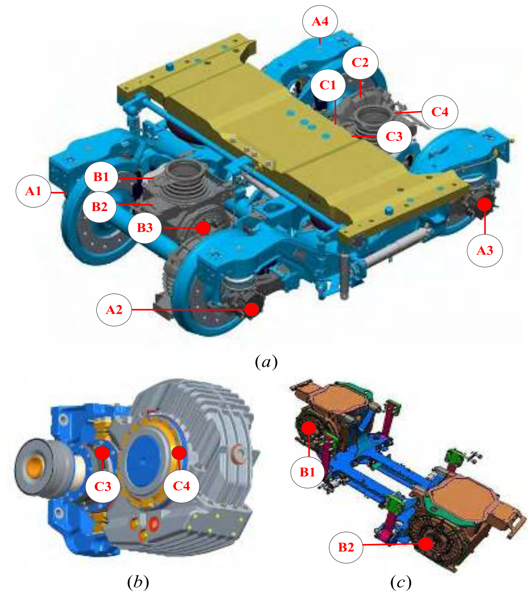


FIGURE 1. Structure of the running gears of a high-speed train.

II. PRELIMINARIES AND PROBLEM FORMULATION

The running gears of a high-speed train are a multi-level complex structure composed of multiple mechanical components. Figure 1 comprises 3 sub-figures: (a) running gear, (b) gear box, (c) traction motor, and other components. Therefore, it is very difficult to establish a complete fault diagnosis model based on the analytical model. In addition, there are a large number of sensor arrangements for the running gears of a high-speed train. For example, consider the temperature sensor measuring point in figure 1, A1–A4: axle box bearing temperature measuring point, B1–B3: motor temperature measuring point, and C1–C4: gear box temperature measuring point. It is easier to get a lot of monitoring data at run-time. However, the running gears of a high-speed train are highly reliable, and they are regularly maintained. It is difficult to acquire a large amount of failure data since the data-driven method has limitations on the training of the model. This paper proposes a semi-quantitative information model in the diagnosis of the fault status of the running gears of a high-speed train, which mainly solves the following three problems.

Problem I. Many features of the running gears of a high-speed train can describe the fault diagnosis of the system such as wheel pair wear, axle box vibration, and gear box temperature being too high. A fault diagnosis model built with multiple features yields more accurate results. However, the complexity of building a model and its calculation

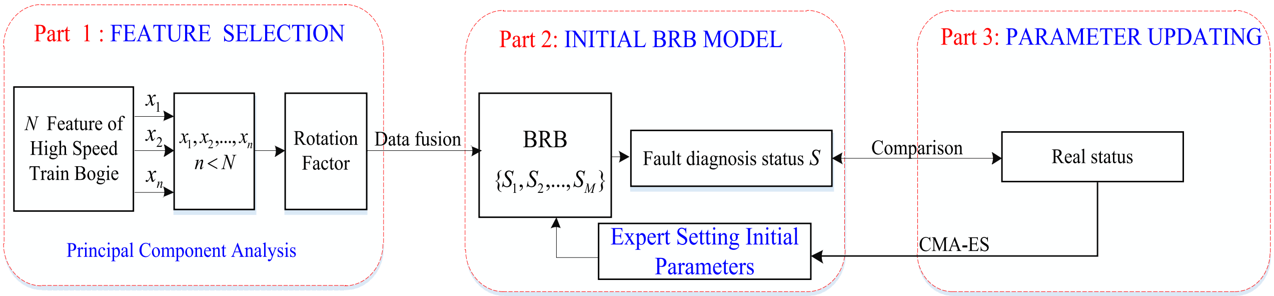


FIGURE 2. **The structure of semi-quantitative information algorithm.**

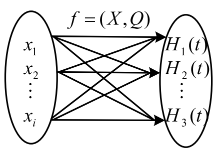
exponentially increases using numerous characteristic features. According to the mechanism analysis, the key feature that can reflect the state of running gears of a high-speed train is denoted by

$$\{x_1, x_2, \cdots, x_i, \cdots, x_n\} \rightarrow \{x_1, x_2, \cdots, x_i\} \tag{1}$$

where x_i denotes a key feature that reflects the state of the running gears of a high-speed train.

Therefore, Problem I deals with how to screen all the key features and ensures the physical meaning of embedding qualitative expert knowledge.

Problem II: Based on the screening of key features, a nonlinear model with the fusion of expert knowledge and sample data for running gears of a high-speed train is designed,



where $H(t) = \{(S_m, \phi_m), m = 1, 2, \cdots, M\}$ denotes the status at time t, S denotes the status of the running gears of high-speed train, and ϕ denotes the belief level for the estimation of S_m,$X = \{x_1, x_2, \cdots, x_i\}$ denotes the monitoring data for the key features, $f = (X, Q)$ denotes the established nonlinear system function, Q denotes the parameter for the nonlinear model.

Thus, Problem II is how to make use of prior knowledge and effective monitoring data to build a nonlinear fault diagnosis model.

Problem III: The starting parameter Q needs to be set in advance based on historical information and the expert knowledge in the initial BRB model. The initial parameter, Q is an incomplete range because of the limited expert knowledge, and thus, the initial value of Q is inaccurate.

Therefore, Problem III deals with how to obtain an accurate fault diagnosis model of the running gears of a high-speed train by optimizing the parameter Q.

III. A SEMI-QUANTITATIVE INFORMATION BASED FAULT DIAGNOSIS METHOD FOR THE RUNNING GEARS SYSTEM OF HIGH-SPEED TRAINS

To solve these three problems, figure 2 shows a semi-quantitative information model to estimate the running gears system. Theof high-speed trains. This method comprises three parts. First part deals with the selection of the key features as input and the rotation factors are used to give the key features physical meaning. The second part considers a nonlinear BRB model based on the fusion of expert knowledge and sample data. In the third part, the model is compared to the real status and CMA-ES is used to adjust the parameters.

A. PRINCIPAL COMPONENT ANALYSIS FEATURES SELECTION

The number of features of the running gears of high-speed trains reflecting its status is large. When these features are selected as input, the BRB model will be delayed and the diagnosis system may even crash. Therefore, this selection is necessary to assist the BRB model for key features. The PCA is one of the most effective features of extraction dimension reduction methods. To achieve a good processing effect, the problem is analyzed in a clear way, and the data of the running gears of a high-speed train is divided into four steps:

Step 1: This step establishes a running gear of a high-speed train data feature quantity sample matrix, and arranges all the samples in the center, here N features are assumed, i.e.,

$$\hat{x_i} = x_i - \frac{1}{N}\sum_{j=1}^{N} x_j \tag{2}$$

$$X = \{\hat{x_1}, \hat{x_2}, \cdots, \hat{x_N}\} \tag{3}$$

where $\hat{x_i}$ is the column vector and each column denotes a feature of the running gear of a high-speed train.

Step 2: In this step, we calculate the covariance matrix of the running gear of a high-speed train data, where each element is the covariance between the different components and X. This forms a covariance matrix, which is a real symmetric matrix, i.e.,

$$C_{n\times n} = \begin{pmatrix} c_{11} & c_{12} & \cdots & c_{1n} \\ c_{21} & c_{22} & \cdots & c_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ c_{n1} & c_{n2} & \cdots & c_{nn} \end{pmatrix} \tag{4}$$

where, $c_{i,j} = cov(Dim_i, Dim_j)$.

Step 3: The eigenvalues and eigenvectors of the covariance matrix are obtained. According to the nature of real symmetric matrices, we obtain the following

$$C_{n\times n} = P\Lambda P^{-1} = (\chi_1\chi_2\cdots\chi_n)\begin{pmatrix}\lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n\end{pmatrix}(\chi_1\chi_2\cdots\chi_n)^{-1} \tag{5}$$

where $\lambda_i(i=1,2,\cdots,n)$ is the eigenvalue of the covariance matrix and $\chi_i(i=1,2,\cdots,n)$ is the eigenvector corresponding to λ_i.

Step 4: The running gear of a high-speed train feature values are sorted in descending order, and the number of principal elements and corresponding feature vectors are obtained according to the principal element contribution rate.

We aim at a large amount of data monitored using the running gears of a high-speed train. Based on the above four steps, this study uses the "maximum balance value" factor to ensure the physical meaning of its data. According to the principles of "cumulative principal component greater than 0.70" and "eigenvalue greater than 1," the key features of the principal component contribution are selected.

B. A SEMI-QUANTITATIVE INFORMATION BRB METHOD OF FAULT DIAGNOSIS

The system for the running gears of a high-speed train comprises many belief rules of BRB model, which can be described as follows

$$\begin{aligned} R_k: & \textit{If}\quad x_1 \textit{ is } A_1^k \bigwedge x_2 \textit{ is } A_2^k \cdots \bigwedge x_M \textit{ is } A_M^k \\ & \textit{Then}\{(P_1,\phi_{1,k}),\cdots,(P_N,\phi_{N,k})\} \\ & \textit{With a rule weight } \alpha_k \\ & \textit{and attribute weight } \eta_{1,k},\eta_{2,k},\cdots,\eta_{M,k} \end{aligned} \tag{6}$$

where $x_i(i=1,2,\cdots,M)$ is the i-th antecedent attribute; $A_i^k(i=1,2,\cdots,M;k=1,2,\cdots,L)$ is the value of the x_i-th antecedent attribute in the k-th rule; L is the total number of rules in the BRB; $\phi_{j,k}$ is the belief degree of conclusion part in the k-th rule for the j-th estimated result P_j; α_k is the rule weight of k-th rule; and $\eta_{i,k}$ is the weight of the i-th antecedent attribute.

The results of belief degree can be calculated by the evidence reasoning (ER) algorithm. The final expected utility of the running gears of high-speed train is obtained by the following three steps.

Step 1: We calculate the antecedent attribute matching degree $\varphi_{i,j}^k$, which is the degree of matching between the antecedent attribute and the rule, i.e.,

$$\varphi_{i,j}^k = \begin{cases} 1-\varphi_{i,j}^k, & k=l+1; \\ \dfrac{A_i^{l+1}-x_i}{A_i^{l+1}-A_i^l}, & k=l(A_i^l \le x_i \le A_i^{l+1}); \\ 0, & k=1,2,\cdots,N(k\neq l,l+1). \end{cases} \tag{7}$$

Step 2: The belief rules activation weight ψ_k is calculated in this step, and the antecedent attribute can activate some belief rules in the BRB model. Thus, we have that

$$\psi_k = \frac{\alpha_k \prod_{i=1}^M(\varphi_i^k)^{\overline{\eta_i}}}{\sum_{l=1}^L \alpha_l \prod_{i=1}^M(\varphi_i^l)^{\overline{\eta_i}}} \tag{8}$$

$$\overline{\eta_i} = \frac{\eta_i}{\max\limits_{i=1,2,\cdots,M}\{\eta_i\}} \tag{9}$$

where $\psi_k \in [0,1]$, $k=1,2,\cdots,l$, $\overline{\eta_i}$ denote attribute weights.

Step 3: Using the ER algorithm, the final output $S(x)$ of the BRB model obtained by

$$S(x) = (P_j,\hat{\phi}_j), \quad j=1,2,\cdots,N \tag{10}$$

where $\hat{\phi}_j$ is relative to the belief degree of estimated results P_j. Thus, we have that

$$\begin{aligned} \hat{\phi}_j &= \frac{\mu\times[\prod_{k=l}^L(\psi_k\phi_{j,k}+\rho)-\prod_{k=l}^L\rho]}{1-\mu\times[\prod_{k=1}^L(1-\psi_k)]} \\ \mu &= [\sum_{j=1}^N\prod_{k=l}^L(\psi_k\phi_{j,k}+\rho)-(N-1)\prod_{k=1}^L\rho]^{-1} \\ \rho &= 1-\psi_k\sum_{i=1}^N\phi_{i,k} \end{aligned} \tag{11}$$

where ψ_k can be calculated using equation (8); $\hat{\phi}_i$ is a function of the rule weight α_k $(k=1,2,\cdots,L)$, the belief degree $\phi_{j.k}$ $(j=1,2,\cdots,N;k=1,2,\cdots,L)$ and $\overline{\eta_i}$ $(i=1,2,\cdots,L)$ is the attribute weight.

The expected utility of the final running gear of a high-speed train $S(x)$, from the above three steps, can be obtained by

$$p = \mu(S(x)) = \sum_{j=1}^N \mu(P_j)\phi(j) \tag{12}$$

The expectant utility p of the running gears of a high-speed train can reflect its status.

C. IMPROVED PARAMETERS OPTIMIZATION ALGORITHM FOR CMA-ES

The expected utility p of the running gear of high-speed train system can be calculated using equation (12) to reflect its status. Then, the following objective function $\beta(Q)$ is established, i.e.,

$$\beta(Q) = \frac{1}{U}\sum_{n=1}^U(\tilde{p}_n-p_n)^2 \tag{13}$$

where $Q=[\alpha_k,\eta_i,\phi_{j,k},\mu(P_j)]^T$ is the column vector of the BRB parameters, U is the number of data of key feature, η_i, $\phi_{i,j}\mu(P_j)$ is given by equations (8), (9) and (12). The $\tilde{p_n}$ and p_n values are similar, and the constructed objective function takes the minimum value $\min\limits_Q\{\beta(Q)\}$.

The objective function constraints are as follows:

$$\sum_{j=1}^{M}\phi_{j,k} \leq 1, \quad k = 1, 2, \cdots, L$$
$$0 \leq \alpha_k \leq 1, \quad k = 1, 2, \cdots, L$$
$$0 \leq \bar{\eta}_i \leq 1, \quad i = 1, 2, \cdots, N$$
$$0 \leq \phi_{j,k} \leq 1, \quad j = 1, 2, \cdots, M \tag{14}$$

where α_k and $\phi_{j,k}$ can be obtained using the BRB model (6) and denote the estimated attribute weights.

The objective function problem under the constraint of formula (14) is solved using the CMA-ES algorithm in this paper. The CMA-ES algorithm controls the optimization direction of the whole parameters by controlling the covariance matrix and finds a fast convergence of small population to get the optimal solution. It also effectively solves the unconstrained optimization problems and boundary constraints. Therefore, we introduce the CMA-ES algorithm in the BRB model using the following four steps.

Step 1: We use the number of samples as expectations and generate a population with a normal distribution as shown in equation (15), where Q is the population of initial expectation. We have that

$$Q_q^{g+1} \sim mean^g + \nu^g N(0, C^g)(q = 1, 2, \cdots, \rho) \tag{15}$$

where C denotes the covariance matrix, ν denotes the step size, and g denotes the expected mean of the g-th generation.

Step 2: The constraint condition is transformed into the constraint objective function in this step. The vector parameters of BRB imports the belief degree of the k-th rule, where $\phi_{j,k} \in Q_q$ and $\sum_{j=1}^{N}\phi_{j,k} = 1$. The constraint equation is transformed into the constraint object function by

$$H_k(\phi_{j,k}) = |\sum_{j=1}^{N}\phi_{j,k} - 1| \tag{16}$$

where $H_k(\phi_{j,k})$ denotes the equation constraint object function of k-th rule in BRB.

Step 3: Recombination and selection: The expected mean will be offset to optimize the sample population distribution. We assume that the population evolved to obtain the updated mean, and choose τ as the optimal solution to update the average. Thus, we have that

$$mean^{g+1} = \sum_{i=1}^{\tau}\gamma_i Q_{i,k}^{g+1} \tag{17}$$

where λ denotes the number of offsprings, $Q_{i,k}^{g+1}$ denotes the i-th solution selected from the λ offsprings in the $g+1$ generation, and γ denotes the offspring weight. The sum of the weights is equal to 1.

Step 4: The covariance matrix is obtained using equations (18) to(21). Thus, we have that

$$C^{g+1} = (1 - a_1 - a_\tau)C^g + a_1 b^{g+1}(b^{g+1})^T + a_\tau \sum_{i=1}^{\tau}\gamma_i\left(\frac{Q_{i,k}^{g+1} - mean^g}{\nu^g}\right)\left(\frac{Q_{i,k}^{g+1} - mean^g}{\nu^g}\right)^T \tag{18}$$

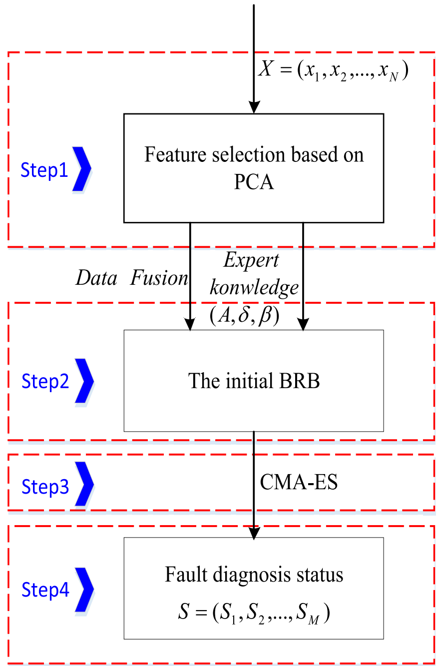


FIGURE 3. Fault diagnosis process of the running gears of a high-speed train.

where a_1 and a_τ denote learning factors and b denotes the evolutionary path. b can be obtained using equation (19), i.e.,

$$b^{g+1} = \sqrt{a_b(2 - a_b)(\sum_{i=1}^{\tau}\gamma_i^2)^{-1}}\frac{mean^{g+1} - mean^g}{\nu^g} + (1 - a_p)b^g \tag{19}$$

where $a_b \leq 1$ denotes a parameter of the evolution path, τ denotes the step size, and ν denotes the update. It can be obtained using equation (20), i.e.,

$$\nu^{g+1} = \nu^g exp(\frac{a_\nu}{d_\nu}(\frac{\|b_\nu^{g+1}\|}{E\|N(0, I)\|})) \tag{20}$$

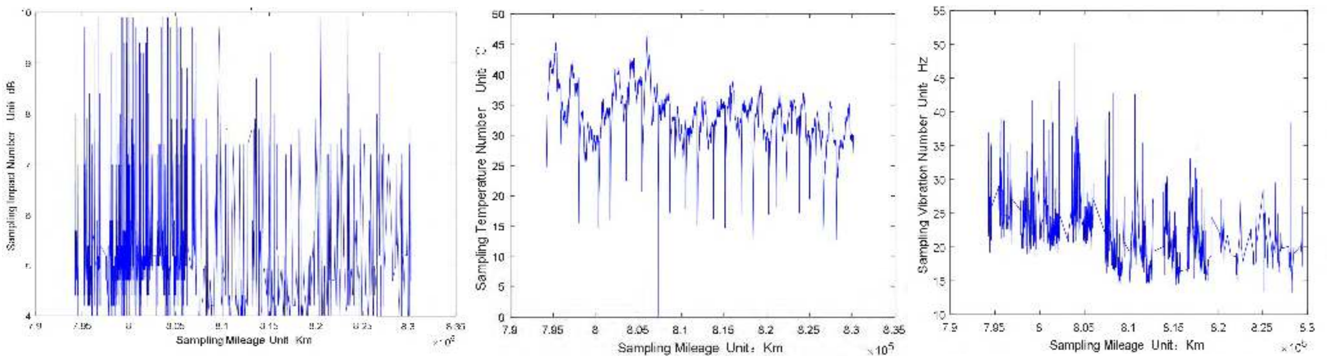
where I denotes the identity matrix, $E\|N(0, I)\|$ denotes the expectation of $\|N(0, I)\|$, d_ν denotes the damping coefficient, a_ν denotes the parameter of evolution path b_ν, and b_ν is updated using equation (21), i.e.,

$$b_\nu^{g+1} = \sqrt{a_\nu(2 - a_\nu)(\sum_{i=1}^{\tau}\gamma_i^2)^{-1}}C^{(g)-\frac{1}{2}}\frac{mean^{g+1} - mean^g}{\nu^g} + (1 - a_\nu)b_\nu^g \tag{21}$$

Using the above iteration process, the optimal parameter Q can be obtained under the premise of satisfying the accuracy.

D. THE STEPS OF FAULT DIAGNOSIS MODEL

Using the above algorithm, the four steps below generalize the fault diagnosis algorithm of the running gears of a high-speed train as shown in figure 3.



(a) (b) (c)

FIGURE 4. Trends of (a) impact, (b) temperature, and (c) vibration of monitoring data.

Step 1: In this step, we select features using PCA in the following ways: a sample matrix is created using equation (3). Then, the covariance matrix (4) is calculated. Furthermore, the eigenvalue is calculated using equation (5), and the key features are screened.

Step 2: We establish a nonlinear parameter BRB model: the BRB model is based on the attribute weight, the reference value, and the belief degree.

Step 3: The CMA-ES algorithm optimizes the parameters in the following ways: we complete the sampling operation (15), perform multi-objective constraints (16), reorganization and selection (17), and update the covariance matrix (18) to (21).

Step 4: The diagnosis of the running gears of a high-speed train can obtain a state of health with the belief degree using the ER algorithm.

The shaft temperature and the gear box vibration are selected to estimate the status of running gears of a high-speed train. If the shaft temperature for the health estimation of the running gears of a high-speed train is same as that of gear box vibration, the attribute weight can be determined to be 1, in which the attribute weight represents the significance of the health status. To measure the change in the evaluation input and to estimate the health status, we assume that the reference values are "normal," "general," and "fault." For example, when the running gears of a high-speed train are in a normal state, the shaft temperature is assumed to be from 25°C to 33°C. The shaft temperature is assumed to be above 40°C when the running gear of a high-speed train is in a fault state. Based on the expert knowledge, the values between 33°C and 40°C can be used as a reference for general health conditions. The more the reference values, the more detailed is the process of describing changes in health status. Section 4 describes the setting process of expert knowledge involving this case.

IV. CASE ANALYSIS

This section discusses running gears of a high-speed train as an example of how to improve its diagnostic estimated capability. Also, it verifies the validity and reliability of the proposed semi-quantitative information model. When the running gear of a high-speed train data is collected, the train is selected as the number 2 car running gear in the month for data sampling. To ensure that the train is in operation, the monitoring data with the rotational speed of $1000r/min$ or more is verified as the method of the model. The monitoring indicators of the running gear include temperature, vibration, and impact. Due to the weight, center of gravity, and suspension parameters of the components complexity of the running gears of a high-speed train, its status can be set to "normal," "general," and "fault."

A. DATA PREPROCESSING

There are some characteristics such as a large amount of data among the actual monitoring data, and many abnormal points and strong environmental noise, which ensure that the data is pre-processed. The impact, temperature, and vibration monitoring data contain numerous duplicate data and outlier data, as shown in figure 4.

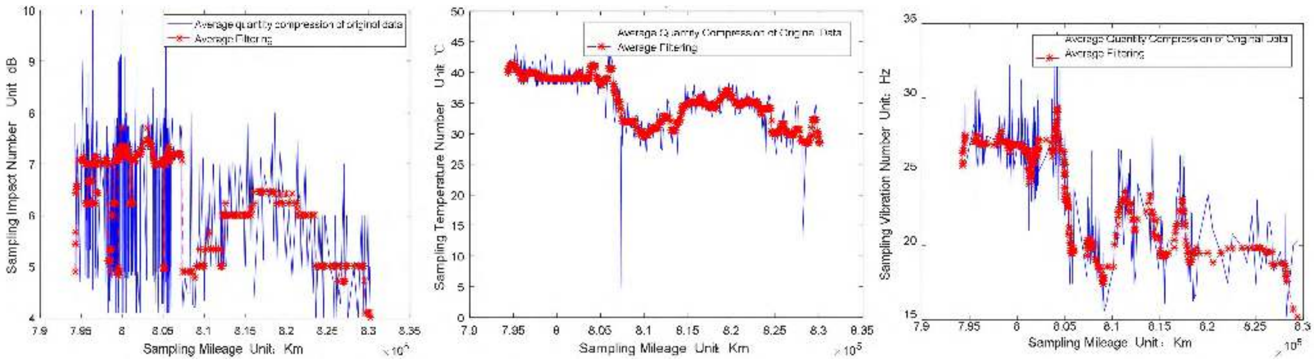
The impact, temperature, and vibration of monitoring data are pre-processed, and the average value is used to reduce the amount of monitoring data on these three points and the average outlier point is filtered using the mean filtering method, simultaneously. Finally, the compressed data volume is 470. The trend graph is shown in figure 5.

B. FEATURES EXTRACTION

The data of three characteristics of temperature, vibration, and impact are dimensionless, and the monitoring data is centralized.

Then, the SPSS software is used to select the key features for PCA. The obtained principal components are shown in Table 1. The relation between the principal components and the key indicators is shown in Table 2.

Based on the principles of "characteristic value greater than 1" and "cumulative principal component greater than 0.70", it is worthy to note that the temperature and vibration are the key feature quantities.



(a) (b) (c)

FIGURE 5. Mean filter trend of (a) impact, (b) temperature, and (c) vibration of monitoring data.

TABLE 1. Initial calculation results calculated using PCA.

Composition	Eigenvalues	Percentage of variance %	Cumulative percentage%
1	1.607	76.555	76.555
2	.877	18.238	94.793
3	.396	5.212	100.000

TABLE 2. Relationship between principal components and key indicators.

Key indicators	Composition 1	Composition 2
Temperature	.889	.121
Vibration	.895	.019
Impact	-.126	.991

C. THE BRB MODEL TRAINING

Considering the analysis of the working principle of the running gears of high-speed train, the status can be divided into the following three states.

Normal state: The status of the running gears of a high-speed train is fine in this case. Under the present circumstances, the temperature and vibration are normal, while the amplitude is low.

General state: Here, the status of running gears of a high-speed train is in between normal state and fault state. In this case, the amplitude value is in the range of 22–26 Hz, while the temperature is in the range of 33°C–36°C.

Fault status: In fault state of the running gears of a high-speed train system, the vibration is severe. The amplitude is up to 35 Hz, and the temperature is up to 46°C.

In the BRB model, not only the vibration and temperature indicators are important for the fault diagnosis of the running gears of a high-speed train, but also the setting of the rules is important; thus, α_k, η_i can be assigned a value of 1. When setting the reference values for vibration and temperature, the number of reference values constitute the number of rules. Therefore, increase in the number of rules will lead to the complexity of calculation of the model.

Based on the expert knowledge, the temperature and vibration reference levels include Normal, Medium, and

TABLE 3. Temperature reference values.

Semantic value	H	U	F
Quantized value	1	2	3

TABLE 4. Vibration reference values.

Semantic value	H	U	F
Quantized value	1	2	3

TABLE 5. Health status reference values.

Semantic value	N	G	F
Quantized value	1	2	3

TABLE 6. Initial BRB parameters setting.

Rule number	Vibration and temperature	Condition distribution. $\{S_1, S_2, S_3\} = \{1, 2, 3\}$.
1	H AND H	$\{(S_1, 0.9), (S_2, 0.1), (S_3, 0)\}$.
2	H AND U	$\{(S_1, 0), (S_2, 0.8), (S_3, 0.2)\}$.
3	H AND F	$\{(S_1, 0.1), (S_2, 0.1), (S_3, 0.8)\}$.
4	U AND H	$\{(S_1, 0.3), (S_2, 0.7), (S_3, 0)\}$.
5	U AND U	$\{(S_1, 0), (S_2, 1), (S_3, 0)\}$.
6	U AND F	$\{(S_1, 0), (S_2, 0.1), (S_3, 0.9)\}$.
7	F AND H	$\{(S_1, 0), (S_2, 0.2), (S_3, 0.8)\}$.
8	F AND U	$\{(S_1, 0), (S_2, 0.1), (S_3, 0.9)\}$.
9	F AND F	$\{(S_1, 0), (S_2, 0), (S_3, 1)\}$.

High expressed as H, U, and F, respectively. Table 3 and Table 4 give the quantified results of temperature and vibration reference levels. The status reference levels include Normal, General, and Fault expressed as H, G, and F, respectively. Table 5 gives the quantified results of status reference levels.

The temperature and vibration reference values can be classified into three levels with a total of nine belief rules. Based on expert knowledge, the fault diagnosis method of running gears of a high-speed train can be established. The BRB model can be given by

$$\begin{aligned} R_k &: \textit{If} \quad \textit{Temperatureis } A_1^k \bigwedge \textit{Vibration is } A_2^k \\ &\textit{Then Status is}\{(1, \phi_{1,k}), (2, \phi_{2,k}), (3, \phi_{3,k})\} \\ &(\sum_{i=1}^{N} \phi_{i,k} \leq 1)\ k \in 1, 2, \cdots, 9 \end{aligned} \tag{22}$$

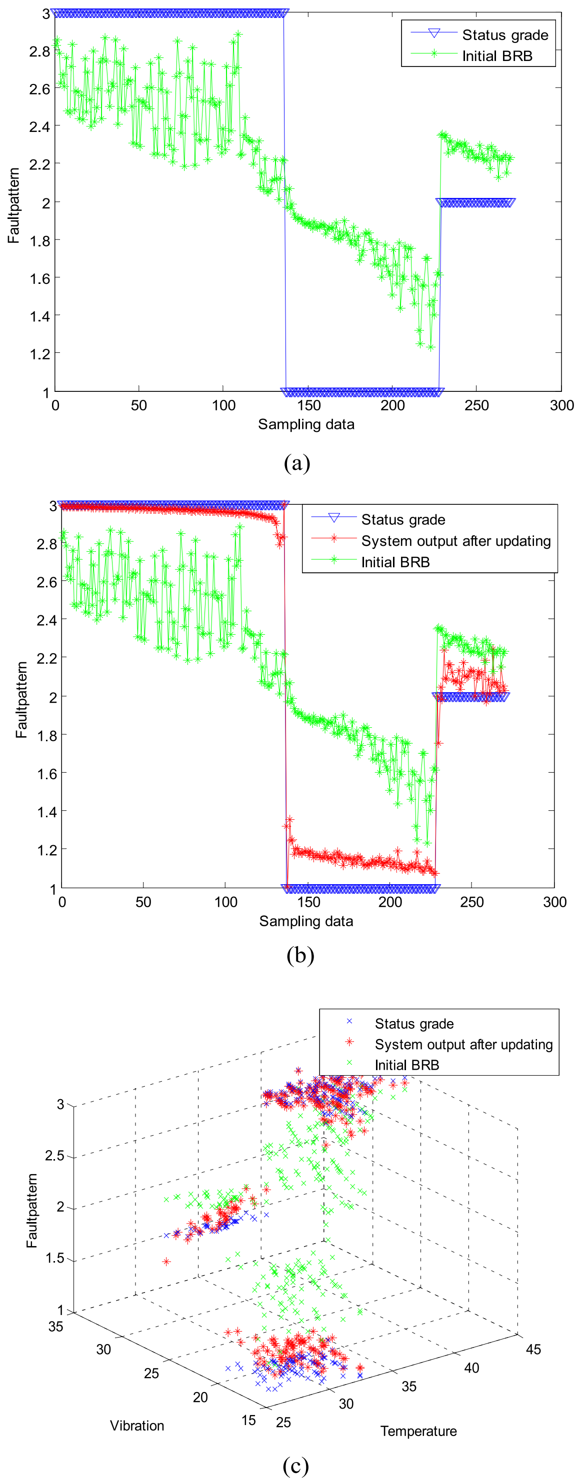


FIGURE 6. **Fault state estimates from initial BRB and updated BRB. (a) The results of initial BRB. (b) The results of updated BRB. (c) Distribution of temperature and vibration data.**

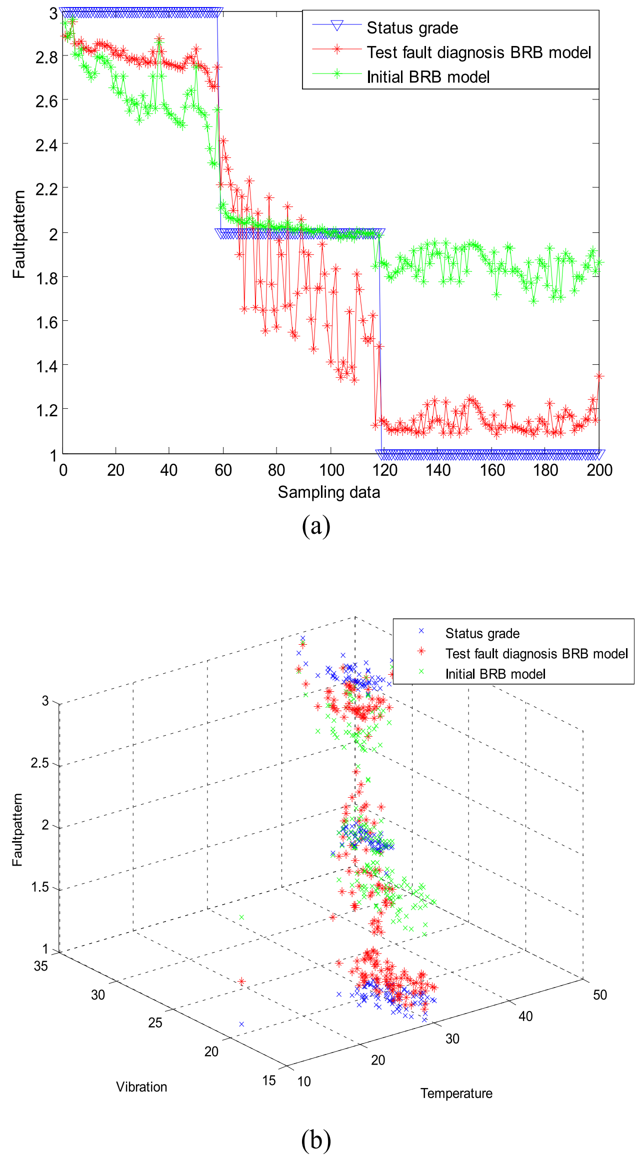


FIGURE 7. **The tested BRB model. (a) The BRB test results. (b) Distribution of temperature and vibration data.**

Based on expert knowledge, Table 6 gives the parameters of the initial BRB model. For example, if the vibration amplitude is large and the temperature is high, the health condition of the running gears of a high-speed train will be very poor. In the first rule H AND H, it means that the temperature and vibration is in a normal state, and the health condition of the running gears of a high-speed train should also be kept in a normal state and is assigned $\{(S_1, 0.9), (S_2, 0.1), (S_3, 0)\}$. The fifth rule explains a gradual process of the running gears of a high-speed train from the normal state to the under-fault state. The rule U AND U indicates that the temperature and vibration are reduced to the general state, and the health condition of running gears of a high-speed train is in a general state. The health condition of running gears of a high-speed train is given by $\{(S_1, 0), (S_2, 1), (S_3, 0)\}$. Figure 6 (a) shows the initial BRB results. The estimated value is not well adapted to the monitoring data, as shown in figure 6 (a), so it is necessary to optimize the initial set parameter values using CMA-ES optimization method in Table 7.

TABLE 7. CMA-ES optimized BRB parameters.

Rule number	Vibration and temperature	Condition distribution. $\{S_1, S_2, S_3\} = \{1, 2, 3\}$.
1	H AND H	$\{(S_1, 0.9102), (S_2, 0), (S_3, 0.0898)\}$.
2	H AND U	$\{(S_1, 0.5699), (S_2, 0.0701), (S_3, 0.3600)\}$
3	H AND F	$\{(S_1, 0.4327), (S_2, 0.3464), (S_3, 0.2209)\}$
4	U AND H	$\{(S_1, 0.9991), (S_2, 0), (S_3, 0.0009)\}$.
5	U AND U	$\{(S_1, 0.1656), (S_2, 0.0569), (S_3, 0.7776)\}$
6	U AND F	$\{(S_1, 0), (S_2, 0), (S_3, 1)\}$.
7	F AND H	$\{(S_1, 0.0307), (S_2, 0.7548), (S_3, 0.2146)\}$
8	F AND U	$\{(S_1, 0.0379), (S_2, 0), (S_3, 0.9620)\}$.
9	F AND F	$\{(S_1, 0), (S_2, 0), (S_3, 1)\}$.

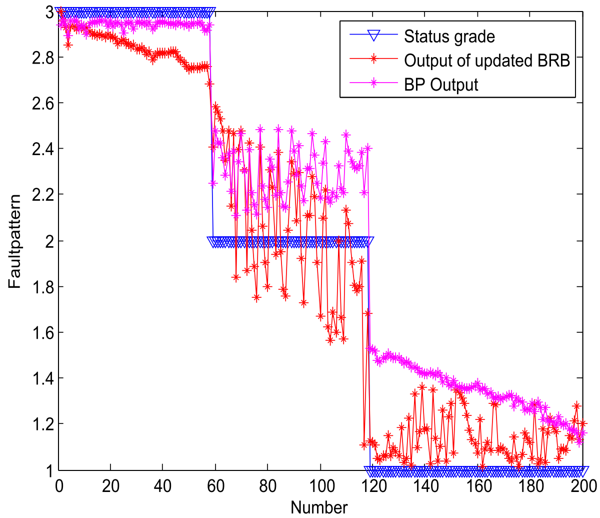


FIGURE 8. The comparative BP neural network.

D. THE BRB TEST MODEL

To further validate the model, the fault state of running gears of a high-speed train is estimated using the back propagation (BP) neural network and SVM model, which are the most classic models in data-driven methods. Differences between describing methods are illustrated by the histogram in Fig. 11, and Table 8 gives the sum of squared residuals for the four different models.

E. COMPARATIVE ANALYSIS

To further verify the model, the fault condition of running gears system is estimate by using the BP neural network and SVM model. BP neural network and SVM model are the most classic models in data-driven methods. Describe the degree of fit of the curve, and Table 8 gives the sum of squared residuals for the five different models.

As shown in figure 8, the output of BP neural network is highly fitted in a healthy normal state, whereas the degree of fit in the general state and the fault state is very low. The fitting SSR value of the BP neural network is 12.09.

The output of the SVM model can fit the monitoring data, although the classification effect is very bad, as depicted in figure 9. The SSR value of the SVM model is 36.00.

As can be seen in figure 10. Though the output of the particle filtering model can fit the monitoring data, the degree of fit in the general state and the fault state is very low. The SSR value of the PF model is 23.11.

TABLE 8. The sum of squared residuals.

Models	SSR
The updated BRB	9.23
Deep learning	62.00
BP Neural Network	12.09
Support Vector Machine	36.00
Particle filtering	23.11

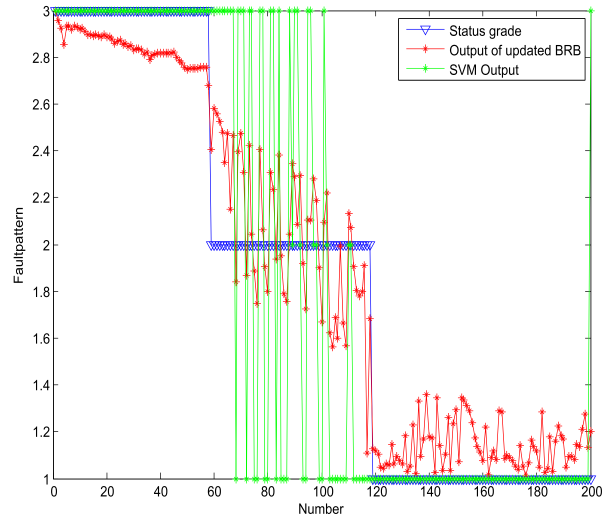


FIGURE 9. The comparative SVM.

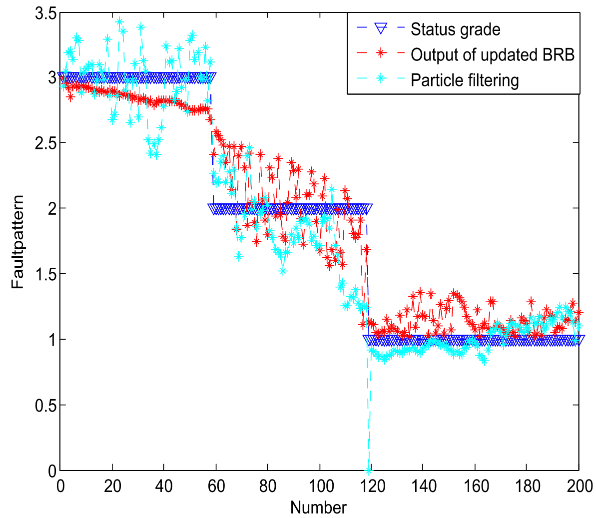


FIGURE 10. The comparative particle filtering.

As can be seen in figure 11. The classification effect of fault state and health state is more accurate, but it is easy to produce the result of misjudgement under the general condition. The SSR value of the deep learning model is 62.00.

Figure 12 is a histogram showing the sum of the squared residuals of the four methods, which can more directly show the advantages of the reliability and effectiveness of the semi-quantitative method compared with other methods.

Therefore, the fault state of the running gears of a high-speed train can be accurately estimated using the

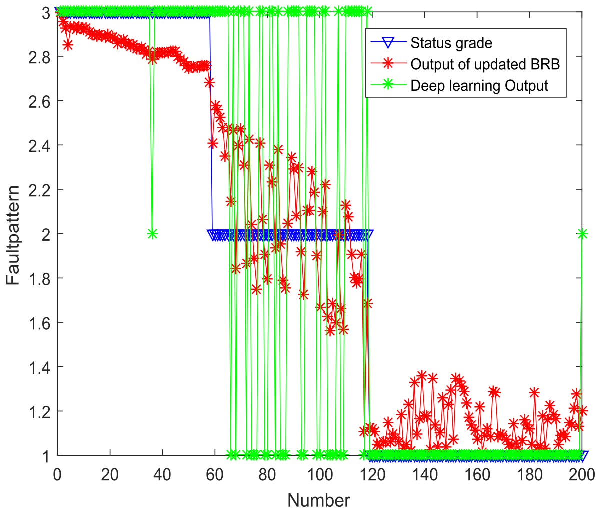


FIGURE 11. The comparative deep learning.

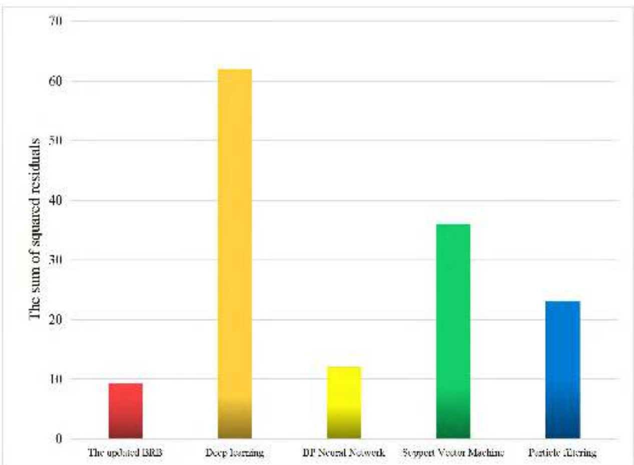


FIGURE 12. The sum of squared residuals for the four different models.

updated BRB. The comparative analysis shows the advantages of semi-quantitative information method on small sample problems.

V. CONCLUSION

To solve the problem of fault diagnosis in the operation of the running gears of a high-speed train, this paper proposed a fault diagnosis method based on semi-quantitative information model. The PCA model was used to extract the monitoring features of the moving parts. The rotation factor was reversed to give physical meaning to the data. The initial model parameters of BRB was used to construct a fault diagnosis method designed to establish the relation between ''poor'' data and expert knowledge. The CMA-ES algorithm was used to optimize the initial BRB model parameters, and the estimation values of the state belief of the running gears of a high-speed train were generated. According to the experimental analysis, using the optimization model proposed in this paper, the BRB model after training can reflect the real-time status of the traveling system well, and has high applicability to practical engineering problems. Therefore, it is impossible to obtain a large amount of effective monitoring data. The project provides a new solution.

The BP neural network, support vector machine, particle filtering, deep learning and semi-quantitative information model were compared with each other. When the failure modal data are limited, the result showed that the semi-quantitative information method combined with expert knowledge effectively solved the problem of running gears of a high-speed train fault diagnosis with high accuracy.

The proposed model input features were varied without implicit information. In practical engineering, the characteristic indicators describing complex systems have implicit information, such as residual life and robustness. Although this paper used the PCA method to select features, it does not add the association analysis among the implicit information characteristics. Moreover, the structure of the algorithm has high latency in calculating multi-dimensional feature quantities, and there are shortcomings in real-time online diagnosis. Our future study is to establish a relevant feature BRB analysis model with implicit information.

REFERENCES

[1] M. Torabi, S. G. M. Mousavi, and D. Younesian, ''A high accuracy imaging and measurement system for wheel diameter inspection of railroad vehicles,'' *IEEE Trans. Ind. Electron.*, vol. 65, no. 10, pp. 8239–8249, Oct. 2018.

[2] J. H. Won, Y. J. Oh, J. Lee, H. W. Ahn, H. Lee, and Y. Cho, ''A study on the individual control method comparing the lateral displacement control of front wheel and rear wheel of IRWs system,'' *IEEE Trans. Appl. Supercond.*, vol. 28, no. 3, Apr. 2018, Art. no. 0601605.

[3] D. Zhou, H. Ji, X. He, and J. Shang, ''Fault detection and isolation of the brake cylinder system for electric multiple units,'' *IEEE Trans. Control Syst. Technol.*, vol. 26, no. 5, pp. 1744–1757, Sep. 2018.

[4] J. Wang, C. Song, P. Wu, and H. Dai, ''Wheel reprofiling interval optimization based on dynamic behavior evolution for high speed trains,'' *Wear*, vols. 366–367, pp. 316–324, Nov. 2016.

[5] W. Deng, J. Xu, and H. Zhao, ''An improved ant colony optimization algorithm based on hybrid strategies for scheduling problem,'' *IEEE Access*, vol. 7, pp. 20281–20292, 2019. doi: 10.1109/ACCESS.2019.2897580.

[6] T. Gumbricht *et al.*, ''An expert system model for mapping tropical wetlands and peatlands reveals south america as the largest contributor,'' *Global Change Biol.*, vol. 23, no. 9, pp. 3581–3599, Sep. 2017.

[7] Z. Zhou, Z. Feng, C. Hu, F. Zhao, Y. Zhang, and G. Hu, ''Fault detection based on belief rule base with online updating attribute weight,'' in *Proc. 32nd Youth Acad. Ann. Conf. Chinese Assoc. Automat.*, May 2017, pp. 272–276.

[8] H. Chen, B. Jiang, W. Chen, and H. Yi, ''Data-driven detection and diagnosis of incipient faults in electrical drives of high-speed trains,'' *IEEE Trans. Ind. Electron.*, vol. 66, no. 6, pp. 4716–4725, Jun. 2019. doi: 10.1109TIE.2018.2863191.

[9] Q. Jiang and X. Yan, ''Parallel PCA–KPCA for nonlinear process monitoring,'' *Control Eng. Pract.*, vol. 80, pp. 17–25, Nov. 2018.

[10] W. Deng, S. Zhang, H. Zhao, and X. Yang, ''A novel fault diagnosis method based on integrating empirical wavelet transform and fuzzy entropy for motor bearing,'' *IEEE Access*, vol. 6, pp. 35042–35056, 2018. doi: 10.1109/ACCESS.2018.2834540.

[11] A. Shebani and S. Iwnicki, ''Prediction of wheel and rail wear under different contact conditions using artificial neural networks,'' *Wear*, vols. 406–407, pp. 173–184, Jul. 2018. doi: 10.1016/j.wear.2018.01.007.

[12] R. Shrivastava, H. Mahalingam, and N. N. Dutta, ''Application and evaluation of random forest classifier technique for fault detection in bioreactor operation,'' *Chem. Eng. Commun.*, vol. 204, no. 5, pp. 591–598, May 2017.

[13] J. Zhao, H. Qu, J. Zhao, and D. Jiang, ''Towards traffic matrix prediction with LSTM recurrent neural networks,'' *Electron. Lett.*, vol. 54, no. 9, pp. 566–568, May 2018.

[14] X. Liu, C. Gong, B. Liu, S. Li, and Z. Xu, "Hidden markov model based signal characterization for weak light communication," *J. Lightw. Technol.*, vol. 36, no. 9, pp. 1730–1738, May 1, 2018. doi: 10.1109/JLT.2018.2792844.

[15] A. Bakhshali, W.-Y. Chan, A. Rezania, and J. C. Cartledge, "Detection of high baud-rate signals with pattern dependent distortion using hidden Markov modeling," *J. Lightw. Technol.*, vol. 35, no. 13, pp. 2612–2621, Jul. 1, 2017. doi: 10.1109/JLT.2017.2691722.

[16] W. Deng, R. Yao, H. Zhao, X. Yang, and G. Li, "A novel intelligent diagnosis method using optimal LS-SVM with improved PSO algorithm," *Soft Comput.*, vol. 23, no. 7, pp. 2445–2462, Apr. 2017. doi: 10.1007/s00500-017-2940-9.

[17] X. He, Z. Wang, Y. Liu, L. Qin, and D. Zhou, "Fault-tolerant control for an Internet-based three-tank system: Accommodation to sensor bias faults," *IEEE Trans. Ind. Electron.*, vol. 64, no. 3, pp. 2266–2275, Mar. 2017.

[18] Z. Chen, S. X. Ding, T. Peng, C. Yang, and W. Gui, "Fault detection for non-Gaussian processes using generalized canonical correlation analysis and randomized algorithms," *IEEE Trans. Ind. Electron.*, vol. 65, no. 2, pp. 1559–1567, Feb. 2018.

[19] T. R. Tronco, M. Garrich, A. C. César, and M. de. L. Rocha, "Cognitive algorithm using fuzzy reasoning for software-defined optical network," *Photonic Netw. Commun.*, vol. 32, no. 2, pp. 281–292, Oct. 2016.

[20] J. Xi, C. Wang, H. Liu, and L. Wang, "Completely distributed guaranteed-performance consensualization for high-order multiagent systems with switching topologies," *IEEE Trans. Syst., Man, Cybern.: Syst.*, to be pubilshed. doi: 10.1109/TSMC.2018.2852277.

[21] A. Swetapadma and A. Yadav, "Data-mining-based fault during power swing identification in power transmission system," *IET Sci. Meas. Technol.*, vol. 10, no. 2, pp. 130–139, Mar. 2016.

[22] Q. Miao, X. Zhang, Z. Liu, and H. Zhang, "Condition multi-classification and evaluation of system degradation process using an improved support vector machine," *Microelectron. Rel.*, vol. 75, pp. 223–232, Aug. 2017. doi: 10.1016/j.microrel.2017.03.020.

[23] Q. Jiang, X. Yan, and B. Huang, "Performance-driven distributed PCA process monitoring based on fault-relevant variable selection and Bayesian inference," *IEEE Trans. Ind. Electron.*, vol. 63, no. 1, pp. 377–386, Jan. 2015.

[24] D. Huang, K. Lanyan, C. Xiaoyan, Z. Ling, and M. Bo, "Fault diagnosis for the motor drive system of urban transit based on improved Hidden Markov modell," *Microelectron. Rel.*, vol. 82, pp. 179–189, Mar. 2018.

[25] H. Zhao, M. Sun, W. Deng, and X. Yang "A new feature extraction method based on EEMD and multi-scale fuzzy entropy for motor bearing," *Entropy*, vol. 19, no. 1, p. 14, Dec. 2016. doi: 10.3390/e19010014.

[26] J.-B. Yang, J. Liu, J. Wang, H.-S. Sii, and H.-W. Wang, "Belief rule-base inference methodology using the evidential reasoning approach-RIMER," *IEEE Trans. Syst., Man, Cybern. A, Syst. Humans*, vol. 36, no. 2, pp. 266–285, Mar. 2006.

[27] R. A. Aliev, W. Pedrycz, O. H. Huseynov, and S. Z. Eyupoglu, "Approximate reasoning on a basis ofZ-number-valued if-then rules," *IEEE Trans. Fuzzy Syst.*, vol. 25, no. 6, pp. 1589–1600, Dec. 2017.

[28] 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 Trans. Syst., Man, Cybern. Syst.*, vol. 48, no. 6, pp. 954–966, Jun. 2018.

[29] M. S. Tonelli-Neto, J. G. M. S. Decanini, A. D. P. Lotufo, and C. R. Minussi, "Fuzzy based methodologies comparison for high-impedance fault diagnosis in radial distribution feeders," *IET Gener., Transmiss. Distrib.*, vol. 11, no. 6, pp. 1557–1565, Apr. 2017.

CHAO CHENG received the B.Eng. degree from Dalian Minzu University, Dalian, China, in 2007, and the M.Eng. and Ph.D. degrees from Jilin University, Changchun, China, in 2011 and 2014, respectively.

He is currently a Teacher with the Changchun University of Technology, Changchun. He is also a Postdoctoral Fellow in process control engineering with the Department of Automation, Tsinghua University, Beijing, China, in 2018. He is also a Postdoctoral Fellow with the National Engineering Laboratory, CRRC Changchun Railway Vehicles Co., Ltd., China, in 2018. His research interests include dynamic system fault diagnosis and predictive maintenance, wireless sensor networks, artificial intelligence, and data-driven method.

XINYU QIAO received the B.Eng. degree from the Changchun University of Technology, Changchun, China, in 2014. He is currently pursuing the M.Eng. degree in computer science and engineering with the Changchun University of Technology.

His research interests include complex system fault prediction, health status prediction, and belief rule base.

HAO LUO received the B.E. degree in electrical engineering from Xi'an Jiaotong University, China, in 2007, the M.Sc. degree in electrical engineering and information technology from the University of Duisburg-Essen, Germany, in 2012, and the Ph.D. degree from the Institute for Automatic Control and Complex Systems (AKS), University of Duisburg-Essen, Germany, in 2016.

He is currently an Associate Professor with the School of Astronautics, Harbin Institute of Technology. His research interests include model-based and data-driven fault diagnosis, fault-tolerant systems, and their plug-and-play application on industrial systems.

WANXIU TENG received the B.Eng. and M.Eng. degrees from Southwest Jiaotong University, Chengdu, in China, in 2002 and 2010, respectively, where he is currently pursuing the Ph.D. degree in mechanical engineering.

He has been a Party Branch Secretary and the Deputy Director of the National Engineering Laboratory, CRRC Changchun Railway Vehicles Co., Ltd., China, since 2015. His research interests include mechatronics measurement technique and fault diagnosis.

MINGLIANG GAO received the B.Eng. degree from Jilin University, Changchun, China, in 2011, and the M.Eng. and Ph.D. degrees from Jilin University, Changchun, China, in 2015.

He has been a Data Supervisor with the National Engineering Laboratory, CRRC Changchun Railway Vehicles Co., Ltd., China, since 2015. He has been a Postdoctoral Fellow in process control engineering with the Department of Automation, Tsinghua University, since 2018. His research interest includes vehicle track big data analysis.

BANGCHENG ZHANG received the B.Eng. and M.Eng. degrees from the Changchun University of Technology, Changchun, China, in 1995 and 2004, respectively, and the Ph.D. degree from Jilin University, Changchun, in 2011.

He has been an Academic Visitor with Tsinghua University, Beijing, China, in 2007. He is currently a Professor with the Changchun University of Technology. He has published over 20 articles. His research interests include mechatronics measurement technique and fault diagnosis.

XIAOJING YIN received the B.Eng. degree from the Liren College, Yanshan University, Qinhuangdao, China, in 2011, and the M.Eng. degree from the School of Mechatronic Engineering, Changchun University of Technology, Changchun, China, in 2014. She is currently pursuing the Ph.D. degree in mechanical engineering with the Changchun University of Technology.

Her research interests include complex system fault diagnosis and prediction, health estimation, and prognostics.