



A WSN Node Fault Diagnosis Model Based on BRB with Self-Adaptive Quality Factor

Guo-Wen Sun¹, Gang Xiang^{2,3}, Wei He^{1,4,*}, Kai Tang¹, Zi-Yi Wang¹ and Hai-Long Zhu¹

¹School of Computer Science and Information Engineering, Harbin Normal University, Songbei District, Harbin, 150025, China

²Beihang University School of Automation Science and Electrical Engineering, Beijing, 100191, China

³Beijing Aerospace Automatic Control Institute, Beijing, 100074, China

⁴Rocket Force University of Engineering, Xi'an, Shaanxi Province, 710025, China

*Corresponding Author: Wei He. Email: he_w_1980@163.com

Received: 30 August 2022; Accepted: 14 December 2022

Abstract: Wireless sensor networks (WSNs) operate in complex and harsh environments; thus, node faults are inevitable. Therefore, fault diagnosis of the WSNs node is essential. Affected by the harsh working environment of WSNs and wireless data transmission, the data collected by WSNs contain noisy data, leading to unreliable data among the data features extracted during fault diagnosis. To reduce the influence of unreliable data features on fault diagnosis accuracy, this paper proposes a belief rule base (BRB) with a self-adaptive quality factor (BRB-SAQF) fault diagnosis model. First, the data features required for WSN node fault diagnosis are extracted. Second, the quality factors of input attributes are introduced and calculated. Third, the model inference process with an attribute quality factor is designed. Fourth, the projection covariance matrix adaptation evolution strategy (P-CMA-ES) algorithm is used to optimize the model's initial parameters. Finally, the effectiveness of the proposed model is verified by comparing the commonly used fault diagnosis methods for WSN nodes with the BRB method considering static attribute reliability (BRB-Sr). The experimental results show that BRB-SAQF can reduce the influence of unreliable data features. The self-adaptive quality factor calculation method is more reasonable and accurate than the static attribute reliability method.

Keywords: Self-adaptive quality factor; belief rule base; wireless sensor networks; fault diagnosis

1 Introduction

A wireless sensor network (WSN) is a physical system for data collection that is often placed in remote areas to collect data [1], such as deep forests, high altitudes, and underwater environments. Now that WSNs work in harsh environments, as the working hours of sensor nodes increases and the impact of the wireless communication environment, the possibility of node failure increases [2]. To



This work is licensed under a Creative Commons Attribution 4.0 International License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

grasp the WSN's working status in time and ensure the reliability of the collected data, WSN node fault diagnosis is essential [3].

Commonly used WSN node fault diagnosis methods are neural network methods, decision tree methods, and random forest methods [4–7]. Among them, neural network methods are the most widely used. WSN node fault diagnosis requires the extraction of data features. However, the data features are not entirely reliable due to the harsh working environments and the interference of wireless signal transmission. However, the above methods do not consider the impact of unreliable data features on the WSN node fault diagnosis process. These unreliable data may lead to anomalous training of the parameters in the method, reducing fault diagnosis accuracy. Meanwhile, most of the above methods are data-driven, with parameters requiring many uniform fault samples to train the parameters for better diagnostic accuracy. However, in actual industrial production, the number of failure samples is small and unlikely to be uniform, which significantly limits the diagnostic accuracy of the data-driven methods.

This paper proposes a WSN node fault diagnosis method based on a belief rule base (BRB) with a self-adaptive quality factor (BRB-SAQF) to solve the above existing problems. The method has two advantages. First, the concept of the quality factor is introduced to reduce the influence of unreliable data. In addition, because BRB is less dependent on the number of training samples and this method is combined with the parameter setting by expert knowledge, which requires fewer fault samples than the data-driven methods. Therefore, good diagnostic results can also be achieved when the number of fault samples is small. To verify the validity of the BRB-SAQF, it is compared with other approaches, such as artificial neural networks, Gaussian regression processes, support vector machines, decision trees and boosting trees.

Through a case study and comparison with the commonly used fault diagnosis methods, it can be concluded that the BRB-SAQF method can effectively reduce the influence of unreliable data on fault diagnosis accuracy. Compared with the data-driven methods, the BRB-SAQF method can also achieve better diagnosis results with the same data samples, and its effectiveness is proven in this paper.

The rest of the article is then structured as follows. Section 2 introduces the current research status of WSN node fault diagnosis and the advantages of the proposed method proposed in this paper. Section 3 defines the problems encountered during the troubleshooting of WSN nodes, and the model's basic structure is illustrated. Section 4 presents the data feature extraction methods, attribute quality factor calculation, model inference process and parameter optimization. In Section 5, the effectiveness of the fault diagnosis method proposed in this paper is verified through a case study. Section 6 summarizes the fault diagnosis method proposed in this paper, and subsequent research work is discussed.

2 Related Work

Due to the widespread use of WSNs, fault diagnosis and classification of WSN nodes have become research topics for related scholars. For example, Saeed et al. proposed a WSN node fault diagnosis method based on supervised learning and integrated learning schemes for extremely random trees [5]. Noshad et al. used a random forest approach to diagnose faults in WSNs [6]. Swain et al. proposed an automatic fault diagnosis model using a hybrid metaheuristic algorithm to train feedforward neural networks for WSN node fault diagnosis [8]. Gharamaleki et al. determined the failure of a WSN by the variability of neighboring nodes [9]. Mohapatra et al. used neural networks to simulate the process of human immunity to achieve WSN node fault diagnosis [10]. Regin et al. used a convolutional

neural network approach for WSN node fault diagnosis [11]. It is easy to see that the most widely used methods for WSN node fault diagnosis are neural network-based methods [12,13].

However, whether it is a neural network-based method or related to the decision tree method, these fault diagnosis methods have three drawbacks. First, these methods do not consider the impact of unreliable data features on fault diagnosis accuracy. Second, they require many uniform fault samples to train the model parameters. Finally, neural network-based methods have a larger number of parameters, such as weights and biases of neurons, which do not have a specific physical meaning, and the interpretability of these methods is poor.

In 2006, Yang et al. proposed a belief rule base inference method based on an evidential reasoning approach (RIMER) [14]. Essentially, RIMER is an expert system consisting of belief rule base (BRB) and evidential reasoning (ER) rules that allow for a more flexible representation of various types of uncertain information, including ambiguity, randomness, and ignorance [15]. The model's parameters are given by experts based on empirical knowledge and have a specific physical meaning in the model. Thus, BRB has the advantage of small sample training [16]. BRB is widely used in medical diagnosis, health status assessment, troubleshooting, etc., [17–20]. Therefore, He et al. proposed a fault diagnosis method for WSNs based on BRB [21]. However, the method also does not consider the effect of unreliable data features on the fault diagnosis process.

In 2018, a new BRB model with attribute reliability was proposed by Feng et al. [22]. When the concept of attribute reliability was introduced, a new static attribute reliability calculation method was proposed. However, the static attribute reliability calculation method does not apply to input attributes with significant variations in different states.

To reduce the impact of unreliable data features on the fault diagnosis process and to bridge the lack of static attribute reliability calculations, this paper proposes a BRB-SAQF-based node fault diagnosis model for WSNs. First, the concept of the attribute quality factor is introduced to mask part of the unreliable data to calculate the quality factors of each attribute. Second, the self-adaptive quality factor calculation method is rederived based on the static attribute reliability calculation method. Third, because BRB has the advantage of small sample training, the required number of the proposed method's training samples is smaller than that of the neural network methods. Finally, the interpretability of the method is also more robust than that of the neural network methods because the settings of the BRB method parameters are determined by expert knowledge and have a specific physical meaning. The current methods of fault diagnosis and the solutions in this paper are represented in Fig. 1.

3 Problem Formulation and Basic Structure of the Model

In this section, some of the problems encountered in the fault diagnosis process of the WSN node are defined, and the model's basic structure is constructed based on these problems.

3.1 Problem Formulation

Problem 1: Several different data features must be extracted as input attributes for the BRB-SAQF model. The raw node data collected by the WSN can be used to determine whether a node is malfunctioning. However, it is impossible to decide what type of failure occurred in this node. Therefore, different data features must be extracted and used to distinguish the fault type of the node. Extracting data features can be described in Eq. (1).

$$[x_1, x_2, \dots, x_M] = f(X_m, \varphi) \quad (1)$$

where $f()$ represents the extraction process of data features and X_m means the raw data collected by WSN sensors. The letter φ represents the parameters in the data feature extraction process. The vector $[x_1, x_2, \dots, x_M]$ represents M data features that are extracted.

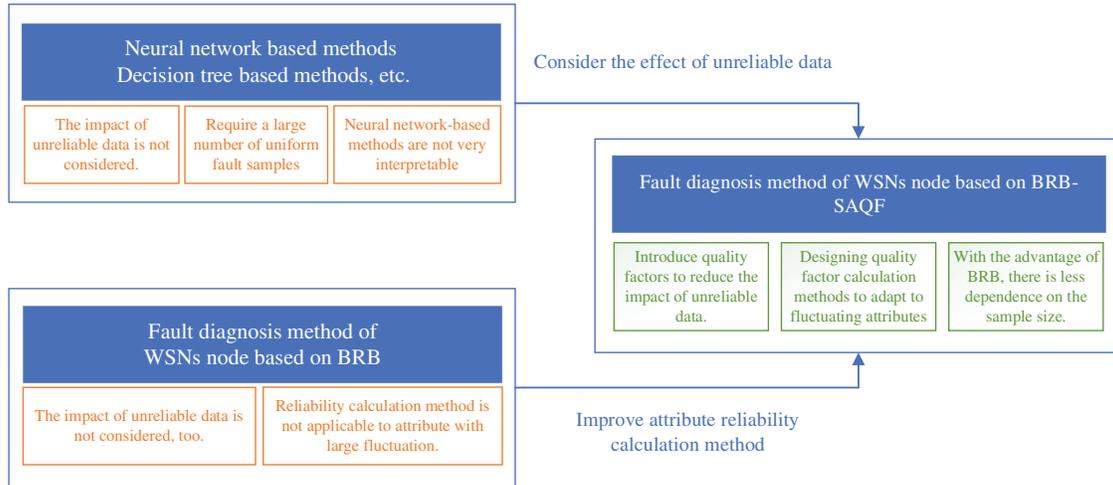


Figure 1: Problems with the current study and solution

Problem 2: The model needs to handle the unreliability of input attributes and define the attribute quality factor parameters. WSN works in a relatively harsh environment, and the data transmission process is often influenced by electromagnetic fields, temperature, humidity and other factors, resulting in noise in the observation data. As a result, the extracted data features contain errors, causing the data features to be incompletely reliable, i.e., the input attributes of the model are unreliable. Therefore, after data feature extraction, it is necessary to identify and deal with the errors caused by noise. The unreliability processing of input attributes can be expressed in Eq. (2).

$$[r_1, r_2, \dots, r_M] = g([x_1, x_2, \dots, x_M], \gamma) \quad (2)$$

where $g()$ represents the calculation function of the attribute quality factor and the vector $[x_1, x_2, \dots, x_M]$ represents the M data features extracted from Problem 1. γ represents the parameters in the attribute quality factor calculation process. The vector $[r_1, r_2, \dots, r_M]$ represents the quality factors of each input attribute.

Problem 3: The attribute quality factor must be incorporated into the BRB inference process. The quality factors of the input attributes are introduced as new parameters for the BRB method. Therefore, it is necessary to consider modifying the BRB's original inference process to consider the attribute quality factors. The BRB inference process incorporating attribute quality factors can be described by Eq. (3).

$$S = h([r_1, r_2, \dots, r_M], \rho) \quad (3)$$

where $[r_1, r_2, \dots, r_M]$ represents the quality factors of the input attributes and ρ denotes other parameters in the fusion process. The symbol $h()$ indicates the BRB-SAQF reasoning process. The alphabet S means the diagnostic result of the model.

Problem 4: The parameters initialized in the BRB-SAQF model must be optimized to obtain more accurate diagnostic results. The optimization process of the model's parameters can be expressed by Eq. (4).

$$\eta_{best} = o(\eta, \psi) \tag{4}$$

where η represents parameters that need to be optimized. The symbol η_{best} represents optimized parameters. The symbol ψ denotes the parameters of the optimization algorithm.

3.2 Basic Structure of the Model

According to the four problems above, the node fault diagnosis model of the WSN based on BRB-SAQF is divided into the following four parts. The first part is the data feature extraction unit, which extracts the data features from the original data that are beneficial to fault diagnosis. The second part is the calculation unit of the model input attribute quality factor, which calculates the attribute quality factor using the self-adaptive quality factor calculation method. The third part is the model rule construction and inference module, which initializes the rules in BRB-SAQF and implements the inference process considering the attribute quality factor. The fourth part is the optimization module of model parameters, which optimizes the initial parameters of the model to obtain better diagnostic results through the optimization algorithm with constraints. The basic structure of the model is shown in Fig. 2.

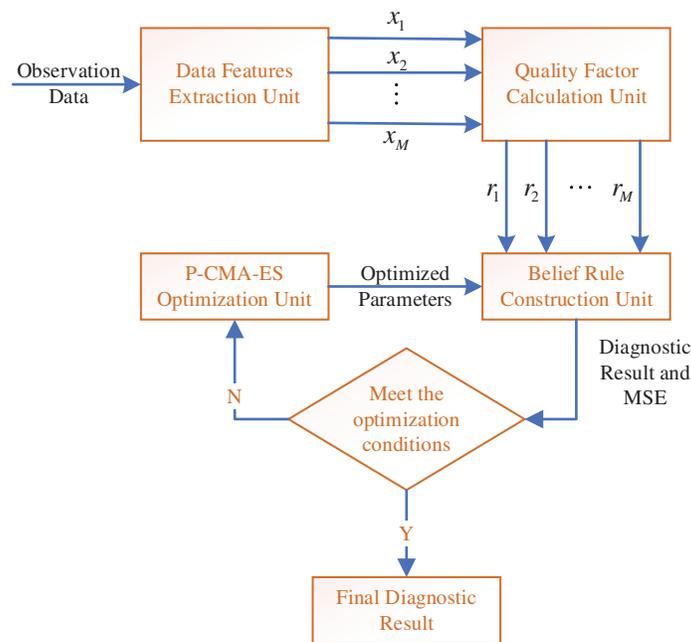


Figure 2: Components of the BRB-SAQF model

4 Construction of the WSN Node Fault Diagnosis Model Based on BRB-SAQF

In Section 3, we defined the critical problems in the node fault diagnosis process of WSNs based on the BRB-SAQF method. In this section, we propose solutions to each of these problems.

4.1 Extraction of Data Features

Different data features must be extracted from the observed data to distinguish what type of failure has occurred on a node. The standard data features used in the WSN node fault diagnosis process are

the mean, standard deviation, variance, skewness and kurtosis [21]. This paper uses mean gap and kurtosis as the extracted data features for the discrimination of different fault types. Both of these data features belong to time-dependent feature data. The mean gap represents the distance between the mean value of the node being diagnosed and the mean value of its neighboring nodes over a period of time. Kurtosis is the characteristic number of the peak height of the probability density distribution curve at the mean [23]. The sensor observation data is defined as $X_m(t)$, which can be described by Eq. (5).

$$X_m(t) = [x_1(t), x_2(t), \dots, x_m(t)] \quad (5)$$

where $x_m(t)$ represents the data collected by m sensors at the moment t . Then, the mean gap from moment l to $l + T$ can be calculated as Eq. (6).

$$g_i = \left| \frac{\sum_{t=l}^{l+T} x_i(t)}{T+1} - \frac{\sum_{t=l}^{l+T} \sum_{j=1}^m x_j(t)}{m(T+1)} \right|, i, j = 1, \dots, m, i \neq j \quad (6)$$

where m denotes the number of proximate sensors. The kurtosis can be calculated by Eq. (7).

$$u_n = \frac{\frac{1}{T+1} \sum_{t=l}^{l+T} [x_i(t) - \bar{x}_i]^4}{\sigma^4} \quad (7)$$

where T indicates the size of the time interval and \bar{x}_i denotes the average value of data collected by the sensor i during the time interval T . The expression σ^4 represents the standard deviation of the data collected by the sensor i during the time interval T .

4.2 Calculating Attribute Quality Factors

When the BRB model considers attribute quality factors, calculating it becomes a new problem. The current research shows that there are several methods to calculate attribute reliability. First, the minimum distance-based attribute reliability calculation method [24] is used. The minimum distance between the data obtained by this method and the accuracy of the attribute reliability calculated by it is not high when there is a lack of observation data [22]. The second method is the expert knowledge-based calculation method [25]. This method is closely related to expert experience. When the experts are not experienced enough or the system is relatively complex, the accuracy of the attribute reliability calculation of this method will be significantly affected [22]. The third is the statistical-based calculation method [26]. This method introduces the concept of the tolerance range, which is used to determine whether the data are reliable and, thus, to calculate the attribute reliability [26]. Therefore, Feng et al. proposed a method to set each attribute's tolerance range and a BRB method based on static attribute reliability (BRB-Sr) [22]. However, this method is suitable for cases where the difference between attributes is slight in different states. When the values of attributes vary widely in different fault states, it is difficult to count and calculate the noise data in the middle segment of attribute values by the static attribute reliability calculation method. Therefore, this paper proposes the self-adaptive attribute quality factor calculation method by improving the static attribute reliability calculation method.

The model assumes that there are M attributes and N diagnostic levels. In this case, the matrix Y can be constructed as shown in Eq. (8).

$$Y = \begin{bmatrix} y_{11} & y_{12} & \dots & y_{1N} \\ y_{21} & y_{22} & \dots & y_{2N} \\ \vdots & \vdots & & \vdots \\ y_{M1} & y_{M2} & \dots & y_{MN} \end{bmatrix} \quad (8)$$

where y_{ij} represents the number of reliable data for the i th attribute corresponding to the j th diagnostic level. The initial value of these elements is 0. Which y_{ij} can be calculated by the following method. Suppose that the samples of the i th attribute corresponding to the j th diagnostic level are $x_{ij}(1), \dots, x_{ij}(k), \dots, x_{ij}(m_{ij})$, $i = 1, \dots, M, j = 1, \dots, N$. The letter m_{ij} is the number of samples, and the total number of samples for the attribute i can be calculated by Eq. (9).

$$m_i = \sum_{j=1}^N m_{ij} \quad (i = 1, \dots, M) \quad (9)$$

Then, the average value \bar{x}_{ij} and the standard deviation σ_{ij} of the m_{ij} samples are calculated. The tolerance range of the i th attribute corresponding to the j th diagnostic level can be described by Eq. (10).

$$[\bar{x}_{ij} - \psi\sigma_{ij}, \bar{x}_{ij} + \psi\sigma_{ij}] \quad (10)$$

where ψ is an adjustment factor used to adjust the size of the tolerance range and is determined by expert knowledge. When data fluctuation is relatively large, a more significant adjustment factor is set, and a smaller adjustment factor is set when the data fluctuation is slight.

Once the tolerance range is determined, it is possible to decide which data are reliable and which are unreliable. If the data presently being judged are within the tolerance range, i.e., $\bar{x}_{ij} - \psi\sigma_{ij} \leq x_{ij} \leq \bar{x}_{ij} + \psi\sigma_{ij}$, then these data are reliable. If the data presently being judged are no longer within the tolerance range, these data are unreliable. When data x_{ij} is reliable, the value of y_{ij} is increased by 1; otherwise, the value of y_{ij} remains unchanged.

At this point, the calculation of element y_{ij} in the matrix Y is completed. The sum of the elements of each row of matrix Y is the number of reliable data for attributes corresponding to the current row. The total number of reliable data for attribute i can be calculated by Eq. (11).

$$y_i = \sum_{j=1}^N y_{ij}, \quad (i = 1, \dots, M) \quad (11)$$

Then, the attribute quality factor can be calculated by Eq. (12).

$$r_i = \frac{y_i}{m_i}, \quad (i = 1, \dots, M) \quad (12)$$

The attribute quality factor calculation process can be summarized in the following four steps and depicted in Fig. 3.

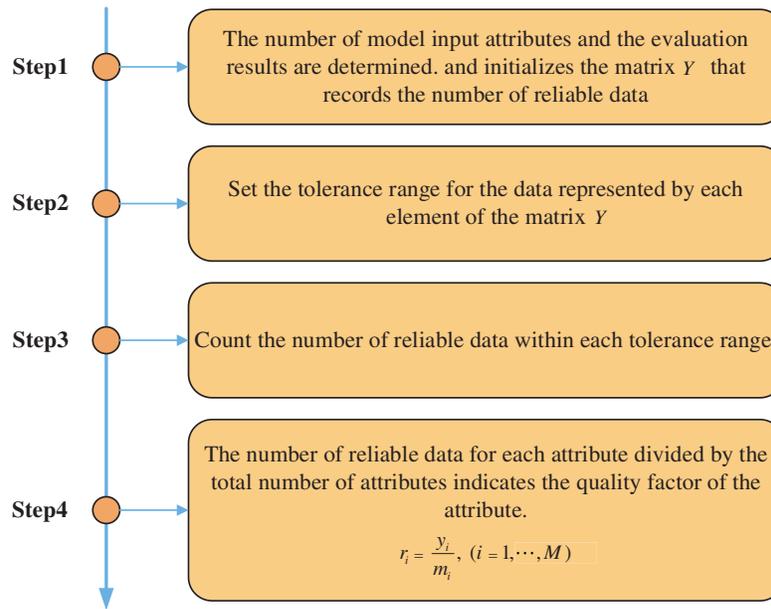


Figure 3: Calculation process of attribute quality factor

4.3 Rule Construction and Reasoning Process of BRB-SAQF

With the preparation of the above two subsections, the data features used as input attributes of the model are extracted, and the quality factor of each attribute is calculated. The next task is to use these data to construct the rules of the BRB-SAQF model. The basic structure of the k th rule R_k is described by Eq. (13).

$$R_k: \text{if } x_1 \text{ is } A_1, x_2 \text{ is } A_2, \dots, x_M \text{ is } A_M$$

$$\text{Then } y \text{ is } \{(D_1, \beta_1), \dots, (D_N, \beta_N)\} \sum_{n=1}^N \beta_n \leq 1 \quad (13)$$

With rule weight θ_k , attribute weight $\delta_1, \delta_2, \dots, \delta_M$ and attribute quality factor r_1, r_2, \dots, r_M

where x_1, x_2, \dots, x_M denotes M input attributes of the model and A_1, A_2, \dots, A_M denotes reference values of M attributes; y is a set of belief distributions described as $\{(D_1, \beta_1), \dots, (D_N, \beta_N)\}$, where D_1, D_2, \dots, D_N are N levels defined by the model and $\beta_1, \beta_2, \dots, \beta_N$ represents the belief degree of each level. θ_k denotes the rule weight, $\delta_1, \delta_2, \dots, \delta_M$ represents the M attribute weight, and r_1, r_2, \dots, r_M represents the M attribute quality factor. Among these parameters of the BRB-SAQF model, rule weight θ_k , attribute weight $\delta_1, \delta_2, \dots, \delta_M$ and belief degree $\beta_1, \beta_2, \dots, \beta_N$ are initialized by expert empirical knowledge.

When the rules of the model are constructed, the next problem to consider is incorporating the attribute quality factor into the reasoning process of BRB. In this paper, the reasoning process of the BRB model considering the quality factor of attributes can be expressed in the following six steps.

Step 1: The attribute matching degree is calculated. When the attribute data are input to the diagnosis model, the first step is to calculate the matching degree of corresponding reference points based on the input values and the reference values of attributes. The calculation process is described by Eq. (14).

$$\alpha_i^j = \begin{cases} \frac{A_i^{k+1} - x_i}{A_i^{k+1} - A_i^k}, & j = k \\ \frac{x_i - A_i^k}{A_i^{k+1} - A_i^k}, & j = k + 1 \\ 0, & j = 1, \dots, K, j \neq k, j \neq k + 1 \end{cases} \quad (14)$$

where α_i^j denotes the matching degree of the i th attribute in the j th reference value, x_i is the current input attribute value, and A_i^k denotes the k th reference value of the i th attribute. When the $A_i^k \leq x_i \leq A_i^{k+1}$ condition is satisfied, the above formula can calculate the matching degree of the attribute reference value.

Step 2: The fusion factor considering the attribute quality factor is calculated. In this step, the attribute quality factor and attribute weight are simultaneously considered and fused into one factor, which is calculated by Eqs. (15) and (16).

$$c_i = \frac{\bar{\delta}_i}{1 + \bar{\delta}_i - r_i} \quad (15)$$

$$\bar{\delta}_i = \frac{\delta_i}{\max\{\delta_i\}} \quad (16)$$

where r_i and δ_i are the attribute quality factor and attribute weight, respectively. The symbol $\bar{\delta}_i$ denotes the relative attribute weight. Eq. (15) shows that when the attribute quality factor is equal to 1, then c_i is equal to 1.

Step 3: The matching degree of the k th rule is calculated. After calculating the attribute matching degree and fusion factor, the matching degree of the rules in the BRB must be calculated. When the matching degree of the rules is not equal to zero, this rule is activated; otherwise, this rule is not activated. The formula for calculating the rule-matching degree is shown in Eq. (17).

$$\alpha_k = \prod_{i=1}^M (\alpha_k^i)^{c_i} \quad (17)$$

where M is the number of attributes in the k th rule and α_k^i denotes the i th attribute matching degree in the k th rule. The symbol c_i is the fusion factor calculated in Step 2.

Step 4: After the rule matching degree is calculated and activated rules are determined, the activation weights of the rules must be calculated. The calculation method of the activation weights is shown in Eq. (18).

$$\omega_k = \frac{\theta_k \alpha_k}{\sum_{i=1}^L \theta_i \alpha_i}, k = 1, \dots, L \quad (18)$$

where θ_k denotes the rule weight of rule k and α_k denotes the rule matching degree. The letter L denotes the number of rules. If the rule is an activated, its activation weight is not equal to 0; otherwise, the activation weight is equal to 0. Note that $0 \leq \omega_k \leq 1$, and $\sum_{k=1}^L \omega_k = 1$.

Step 5: Multiple activated rules are fused, and the fused belief degree can be obtained. After those rules are determined to be activated, they are fused by an evidence reasoning (ER) parsing algorithm [27]. The calculation method is shown in Eqs. (19) and (20).

$$\beta_n = \frac{\mu \left[\prod_{k=1}^L \left(\omega_k \beta_{n,k} + 1 - \omega_k \sum_{j=1}^N \beta_{j,k} \right) - \prod_{k=1}^L \left(1 - \omega_k \sum_{j=1}^N \beta_{j,k} \right) \right]}{1 - \mu \left[\prod_{k=1}^L (1 - \omega_k) \right]} \quad (19)$$

$$\mu = \left[\sum_{n=1}^N \prod_{k=1}^L \left(\omega_k \beta_{n,k} + 1 - \omega_k \sum_{j=1}^N \beta_{j,k} \right) - (N-1) \prod_{k=1}^L \left(1 - \omega_k \sum_{j=1}^N \beta_{j,k} \right) \right]^{-1} \quad (20)$$

where N denotes the fault diagnosis model identification framework with N diagnostic levels and L denotes the number of rules that have been activated. The symbol ω_k denotes the activation weight of the rule and $\beta_{j,k}$ denotes the belief degree of consequence j in the rule k . The initial belief degree is determined by expert knowledge. By calculating Eqs. (19) and (20), the following result shown in Eq. (21) can be obtained.

$$S(x^*) = \{(D_1, \beta_1), \dots, (D_N, \beta_N)\} \quad (21)$$

where x^* represents multiple attributes input into the model, including x_1, x_2, \dots, x_M these M attributes. The symbols D_1, D_2, \dots, D_N represent the identification framework levels of the model. $\beta_1, \beta_2, \dots, \beta_N$ represents the belief degree of each level.

Step 6: The final utility of the model is derived from the results of the fused belief degree. Assume that the utility of the n th level is described by $u(D_n)$. The utility of $S(x^*)$ is obtained by Eq. (22).

$$u(S(x^*)) = \sum_{n=1}^N u(D_n) \beta_n \quad (22)$$

Through the above analysis, the inference process of BRB-SAQF is introduced. The whole inference process can be represented as shown in Fig. 4.

4.4 The Model Optimization Process

In Section 4.3, the initial values of rule weights, attribute weights, and belief degree of BRB-SAQF are determined by expert knowledge. However, when the model has many parameters and the expert has insufficient experience and knowledge, the initial setting of the parameters is not reasonable, which may affect the accuracy of the model diagnosis. Therefore, this paper proposes an optimization process of the model to improve the diagnostic accuracy of the model using a projection covariance matrix adaptation evolution strategy (P-CMA-ES) to optimize the parameters of the model [28,29]. The model's parameters that need to be optimized need to satisfy the following conditions.

For rule weight θ_k , the conditions shown in Eq. (23) need to be satisfied.

$$0 \leq \theta_k \leq 1, k = 1, 2, \dots, L \quad (23)$$

For attribute weight δ_i , the conditions shown in Eq. (24) need to be satisfied.

$$0 \leq \delta_i \leq 1, i = 1, 2, \dots, M \quad (24)$$

For the belief degree of the corresponding consequent in each rule, the conditions shown in Eqs. (25) and (26) need to be satisfied.

$$0 \leq \beta_{n,k} \leq 1, n = 1, 2, \dots, N, k = 1, 2, \dots, L \quad (25)$$

$$\sum_{n=1}^N \beta_{n,k} \leq 1, k = 1, 2, \dots, L \tag{26}$$

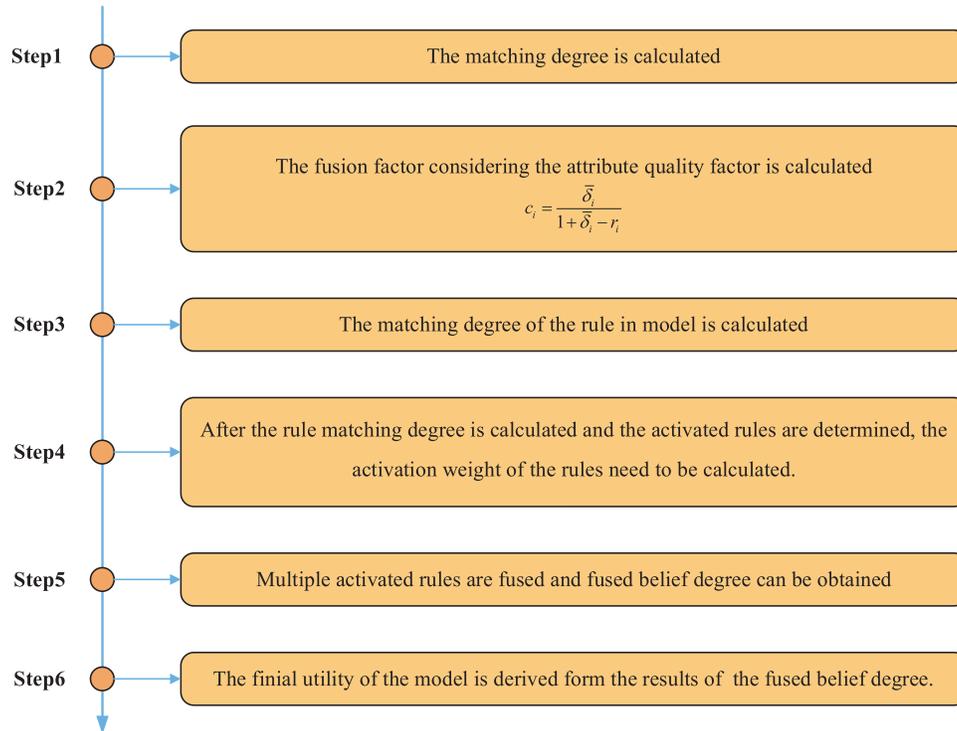


Figure 4: Reasoning process of BRB-SAQF

After the parameters that need to be optimized and the constraints are determined, a metric to reflect the effectiveness of the optimization must be defined. Suppose the diagnosis result of the optimization process is $result_{optimization}$, and the original result of the training data is $result_{origin}$; then, the mean square error (MSE) between the two results is used to reflect the effect of parameter optimization. The MSE is calculated by Eq. (27).

$$MSE(\eta) = \frac{1}{T} \sum_{t=1}^T (result_{optimization} - result_{origin})^2 \tag{27}$$

where T denotes the number of data used to train the model parameters. The symbol η denotes the set of optimized parameters. Through the above analysis, the optimization process of the model can be understood as finding the minimum value of the MSE.

4.5 Model Building Process

By analyzing the contents of this section, the process of building a node fault diagnosis model for WSN based on BRB-SAQF can be divided into the following steps.

Step 1: The data features of the WSN node are extracted from the observed data and used as input attributes of the BRB-SAQF model.

Step 2: The quality factor of the model input attributes is calculated using the self-adaptive attribute quality factor calculation method.

Step 3: The model parameters are initialized, and the model rules are fused using the ER parsing algorithm.

Step 4: The initial parameters are optimized by P-CMA-ES to improve the diagnostic accuracy of the model.

5 Case Study

In this section, we conduct a case study based on sensor data from the Intel Berkeley Research Lab and compare BRB-SAQF with other fault diagnosis methods, including BRB-Sr, artificial neural networks, Gaussian regression, support vector machines, decision trees, and boosting trees. The case study is used to verify that the fault diagnosis method based on BRB-SAQF for the WSN nodes can effectively reduce the impact of noise data on the fault diagnosis process of the WSN nodes.

5.1 Fault Diagnosis Using BRB-SAQF

Step 1: The data in the dataset are preprocessed, and the set of diagnostic results for the model is defined. The dataset consists of temperature, humidity, light and voltage collected by 54 sensors. These sensors are distributed in a laboratory. Then, based on the distribution of the sensors and the trend of the data, the temperature data from sensors 1 to 4 from March 1st to 7th were selected for the study [30]. The dataset was preprocessed because the data were missing at some point. Meanwhile, Gaussian noise was added to the dataset simulation to verify that the model can effectively reduce the interference of noisy data [31,32]. The processed dataset has a total of 2,016 pieces of data with five minutes between two adjacent pieces of data, where the data of sensors 1 to 4 are shown in Fig. 5.

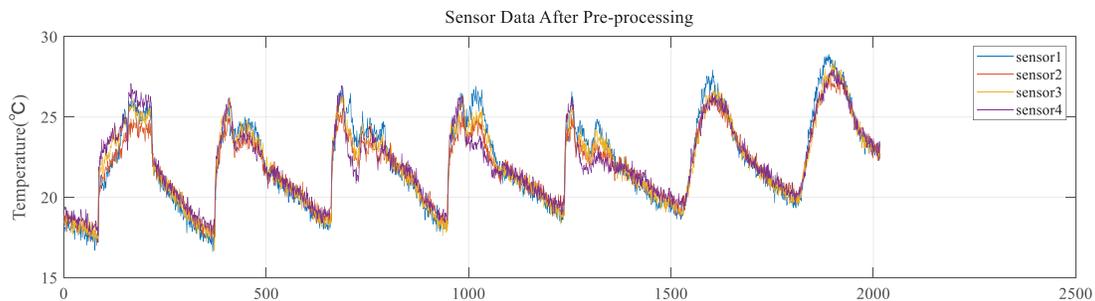


Figure 5: Sensor data after preprocessing

This paper uses offset faults, high-noise faults, outlier faults and fixed value faults as the types of sensor faults to be detected [33]. Next, based on the characteristics of the fault, the four types of faults mentioned above are simulated on sensor 1 using a software approach. The methods of fault data simulation are shown in Table 1, and the fault data after simulation are shown in Fig. 6.

With the above introduction of the case study dataset, the output of the fault diagnosis model identification framework includes normal states (NS), offset faults (OSF), high-noise faults (HNF), outlier faults (OLF) and fixed value faults (FVF), which can be described by Eq. (28), and the corresponding reference values are shown in Eq. (29).

$$\{NS, OSF, HNF, OLF, FVF\} \quad (28)$$

{0, 1, 2, 3, 4} (29)

Table 1: Fault simulation methods

Fault type	Simulation method
Offset faults	The data of sample numbers 400 to 799 are selected by superimposing random numbers in the interval [0, 10].
High noise faults	The data of sample numbers 800 to 1199 are selected by superimposing random numbers in the interval [10, 20].
Outlier faults	10% of the data from sample number 1200 to number 1599 are drawn and replaced with random numbers in the interval [0, 40].
Fixed faults	The collected data is fixed to the value collected before the fault departure.

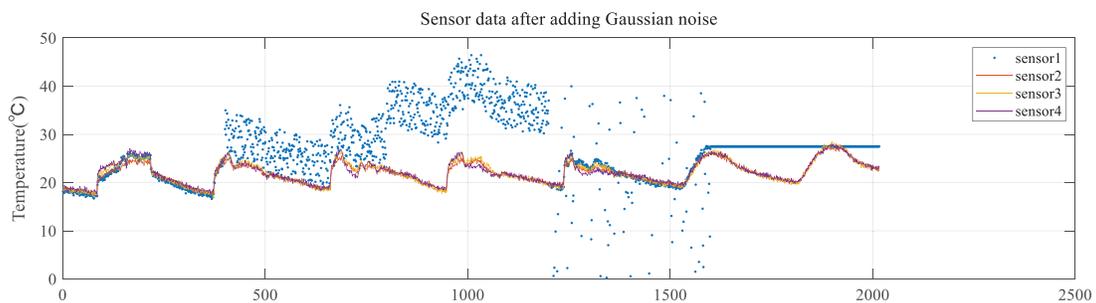


Figure 6: Sensor data after adding Gaussian noise

Step 2: Sensor data features are extracted. After the data are prepared, the data features, including the mean gap and kurtosis, are extracted using the method introduced in Section 4.1 and as input attributes of the model. The time window for extracting data features is set to 12, and the extracted data features are normalized. The data features are shown in Figs. 7 and 8. The highlighted part in the figures displays unreliable data.

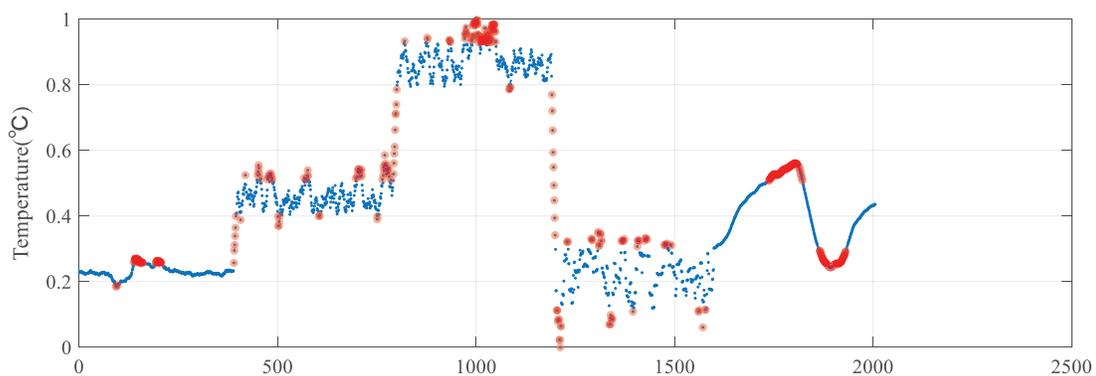


Figure 7: Mean gap data feature image

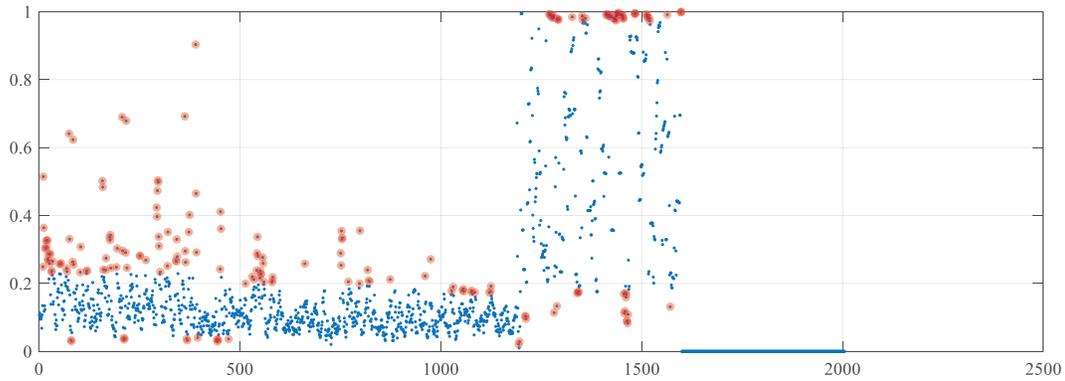


Figure 8: Kurtosis data feature image

Step 3: The reference points and values for the model input attributes are determined. After the data features as input attributes of the model are extracted, each attribute’s reference values and points must be determined. The reference values and points for the mean gap and kurtosis are determined based on the data distribution characteristics in Figs. 7 and 8. The reference points of the mean gap are small (S), relatively small (RS), medium (M), relatively large (RL) and large (L). The reference values of the mean gap are shown in Table 2. The reference points of kurtosis are small (S), relatively small (RS), relatively large (RL) and large (L). The reference values of kurtosis are shown in Table 3.

Table 2: Reference value of mean gap

Reference point	S	RS	M	RL	L
Reference value	-0.001	0.184	0.358	0.700	1.001

Table 3: Reference value of kurtosis

Reference point	S	RS	RL	L
Reference value	-0.001	0.010	0.275	1.001

Step 4: The quality factors of each input attribute are calculated. In this step, the mean gap and kurtosis quality factors are calculated using the method described in Section 4.2. The setting of the tolerance ranges for different cases is shown in Table 4. The quality factors of the mean gap and kurtosis are calculated to be 0.7062 and 0.7276, respectively.

Table 4: Tolerance range parameters of data features

Input attributes	Diagnostic results	Average value	Standard deviation	Adjustment factor
Mean gap	NS	0.2333	0.0213	1.14
	OSF	0.4633	0.0405	1.15
	HNF	0.8723	0.0563	0.94

(Continued)

Table 4: Continued

Input attributes	Diagnostic results	Average value	Standard deviation	Adjustment factor
	OLF	0.2236	0.0702	1.15
	FVF	0.4071	0.0987	0.94
Kurtosis	NS	0.1715	0.1084	0.56
	OSF	0.1085	0.0571	1.02
	HNF	0.1018	0.0519	0.92
	OLF	0.5866	0.2924	0.98
	FVF	0.0090	0.0917	2

Step 5: The other parameters of the model are initialized. These parameters are optimized using the P-CMA-ES method proposed in Section 4.4. After optimization, the parameters are shown in Table 5, where each row represents a rule in the model. Finally, the model inference process presented in Section 4.3 is used together with the optimized parameters to diagnose sensor faults and produce results.

Table 5: Optimized model parameters

Rule number	Input attributes		Rule weight	Belief distribution
	Mean gap	Kurtosis		{NS, OSF, HNF, OLF, FVF}
1	S	S	0.1693	0.21816, 0.09817, 0.27579, 0.13844, 0.26944
2	S	RS	0.4936	0.0024, 0.40539, 0.01229, 0.03631, 0.54362
3	S	RL	0.9784	0.0623, 0.06429, 0.06813, 0.41309, 0.39219
4	S	L	0.9234	0.09458, 0.135282, 0.065474, 0.12399, 0.580676
5	RS	S	0.0062	0.0448, 0.517374, 0.00352, 0.000652, 0.433654
6	RS	RS	0.0118	0.99599, 0.00193, 0.00006, 0.00172, 0.00031
7	RS	RL	0.0028	0.36190, 0.36689, 0.25268, 0.00607, 0.01246
8	RS	L	0.0002	0.30607, 0.07011, 0.12829, 0.10386, 0.39168
9	M	S	0.8700	0.00050, 0.00036, 0.00189, 0.00124, 0.99601
10	M	RS	0.0001	0.63492, 0.05458, 0.09995, 0.15465, 0.05589
11	M	RL	0.0033	0.22425, 0.07192, 0.08152, 0.26334, 0.35897
12	M	L	0.0082	0.02146, 0.17697, 0.07801, 0.11676, 0.60680

(Continued)

Table 5: Continued

Rule number	Input attributes		Rule weight	Belief distribution
	Mean gap	Kurtosis		{NS, OSF, HNF, OLF, FVF}
13	RL	S	0.0094	0.88110, 0.07109, 0.04148, 0.00366, 0.00267
14	RL	RS	0.0723	0.45348, 0.30286, 0.03648, 0.01569, 0.19148
15	RL	RL	0.2681	0.47403, 0.19242, 0.21118, 0.0733, 0.04904
16	RL	L	0.0964	0.21497, 0.18619, 0.08708, 0.38059, 0.13118
17	L	S	0.7690	0.00430,0.09177,0.34379,0.52273,0.03741
18	L	RS	0.9867	0.00018, 0.56913, 0.11641, 0.03250, 0.28178
19	L	RL	0.9347	0.08987, 0.35595, 0.10905, 0.13588, 0.30925
20	L	L	0.3636	0.08792, 0.17882, 0.23044, 0.05537, 0.44745

Step 6: The evaluation indices of the fault diagnosis model are determined. The overall accuracy, false-negative rate (FNR) and false-positive rate (FPR) are used as evaluation indices to verify the method's validity. It is assumed that samples without faults are negative samples and the samples with faults are positive samples. The formulas for the above indices are shown in Eqs. (30)–(32).

$$Accuracy = \frac{Num_{right}}{Num_{sample}} \quad (30)$$

where Num_{right} represents the number of samples correctly diagnosed by the model. The expression Num_{sample} denotes the total number of samples.

$$FNR = \frac{FN}{TP + FN} \quad (31)$$

where FN denotes the number of samples with false negatives and TP denotes the number of samples with true positives.

$$FPR = \frac{FP}{FP + TN} \quad (32)$$

where FP represents the number of samples with false positives and TN represents the number of samples with true negatives.

5.2 Comparison with Other Methods

It is compared with other methods to verify the effectiveness of the proposed method proposed in this paper and the improvement of the BRB-Sr method. These include BRB-Sr, artificial neural networks, Gaussian regression processes, support vector machines, decision trees and boosting trees. The fault diagnosis results compared with the BRB-Sr method are shown in Fig. 9, and the evaluation indices of the different methods are shown in Figs. 10–12.

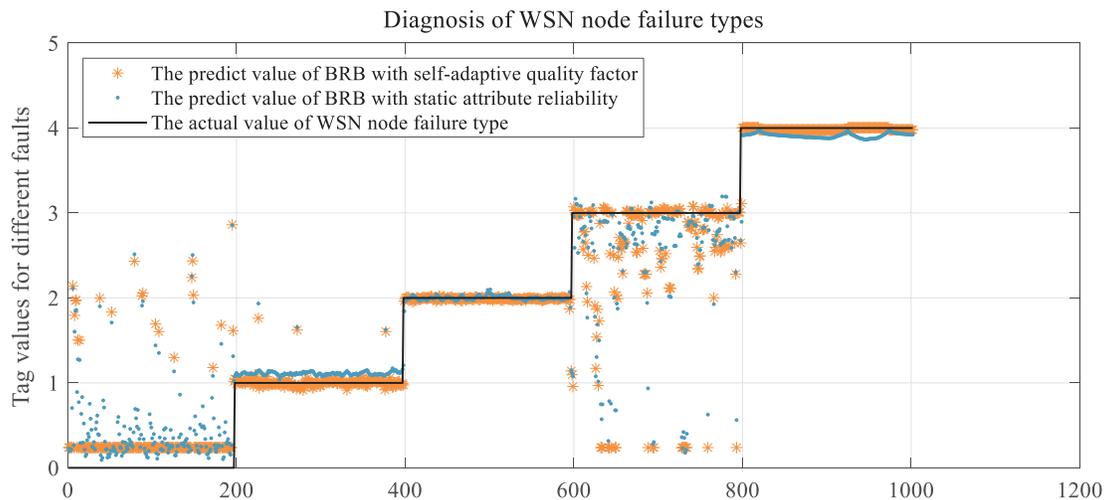


Figure 9: Comparison of diagnostic results of BRB-SAQF and BRB-Sr

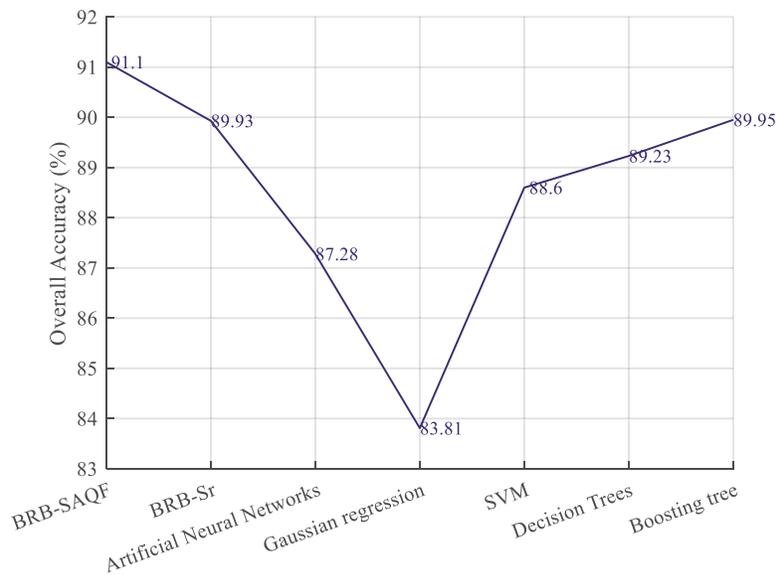


Figure 10: Overall accuracy of different methods

By comparing the diagnostic results of BRB-SAQF and BRB-Sr, it is evident that the value of BRB-SAQF is closer to the actual situation, especially in the reference value of 0. After testing, the overall accuracy, FNR, and FPR of BRB-SAQF are 91.1%, 18.43%, and 1.37%. The BRB-Sr values are 89.93%, 22.94%, and 1.34%, respectively. The reason for the scattering diagnostic results of method BRB-Sr is that the tolerance range of method BRB-Sr cannot wholly count the unreliable data, which leads to the inaccurate calculation of the attribute reliability. The quality factor calculation method obtained by improving the static attribute reliability calculation method can effectively count the unreliable data in the middle region of the fluctuating attribute data, enabling a more focused diagnosis.

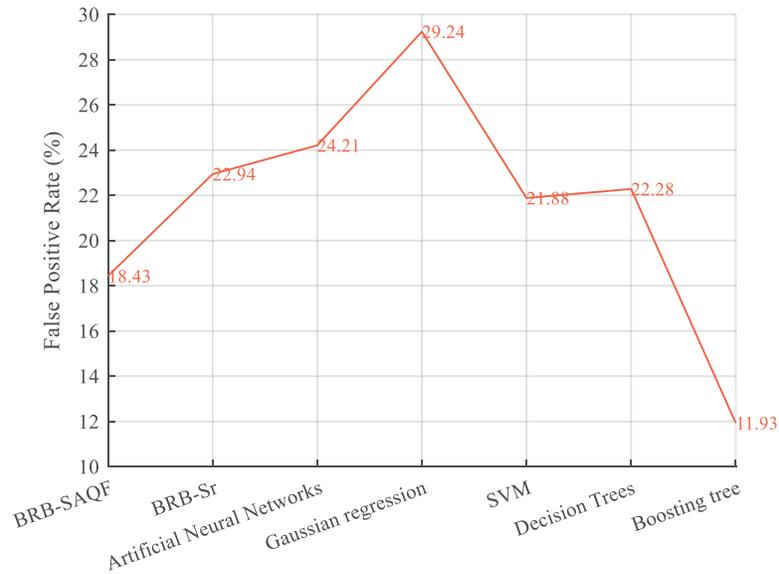


Figure 11: False positive rate of different methods

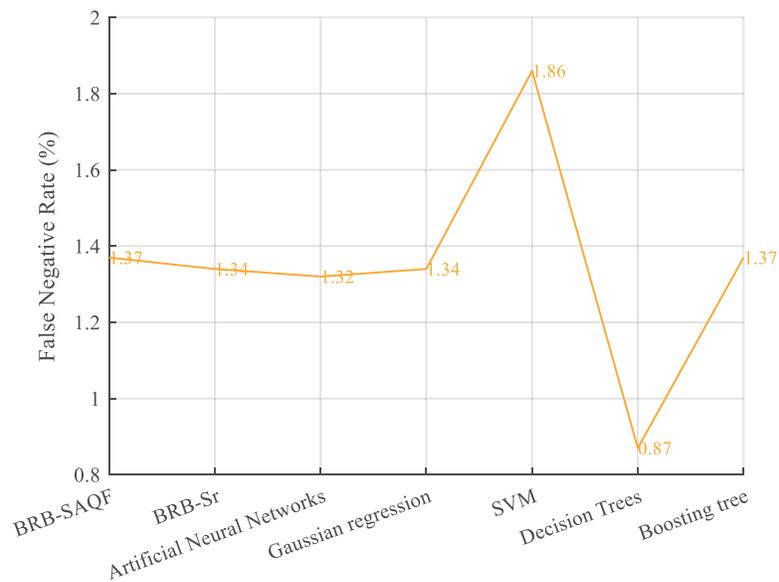


Figure 12: False negative rate of different methods

By comparing the evaluation indices of different methods. First, it can be concluded that the results of the BRB-SAQF and BRB-Sr methods are higher than those of the other methods in terms of overall accuracy. Second, regarding the false-positive rate, the BRB-SAQF method has a relatively low false-positive rate, making it less likely to misdiagnose nodes as faults. Finally, in terms of the false negative rate, the difference between all the above methods is slight, fluctuating at approximately 1.3%, and there is no substantial gap between the methods. The reasons for the above results can be summarized in the following points. First, the BRB-SAQF introduces a quality factor and improves the calculation method of attribute reliability, which can count unreliable data more

effectively. The attribute quality factor calculation is more accurate. Second, the inference process of BRB-SAQF is essentially similar to BRB's, which can handle uncertain information, including ambiguity, randomness and ignorance [15]. Finally, the BRB-SAQF model parameter settings come from expert experience and knowledge, and the parameter settings are more reasonable. Compared with the neural network approach, the BRB-SAQF method is more interpretable and more applicable to training with small, unbalanced samples. With the combined effect of the above factors, the BRB-SAQF method improves the fault diagnosis accuracy of the WSN node.

6 Conclusion

The following conclusions can be obtained through the analysis of related work and the examination of case studies. Some shortcomings exist in the commonly used fault diagnosis methods for WSN nodes. First, the methods do not consider the influence of environmental noise collected by the sensors during the fault diagnosis process. Second, with the weight and bias parameters in the method, the neural network-based methods aim to fit a nonlinear function and require many uniform fault samples to train model parameters. Finally, the attribute reliability calculation renders the BRB-Sr approach unable to count the unreliable data in the middle segment when the attribute value changes considerably, which affects the fault diagnosis accuracy and needs to be optimized to obtain a more accurate diagnostic result.

The BRB-SAQF model is proposed as a fault diagnosis method to solve the above shortcomings. First, the calculation method of the attribute quality factor is designed to compensate for the shortcomings of the static attribute reliability calculation method to improve the accuracy of fault diagnosis. Second, the parameters in the BRB-SAQF method represent the rules, weights and quality factors of the attributes and the probability of possible failures. Experts initialize these parameters based on their experience and knowledge, which are more reasonable and interpretable. Finally, because the BRB method has the advantage of small sample training, the number of fault samples must be smaller. Through the case study in Section 5 of this article, the accuracy of the method proposed in this paper is improved compared with that of the BRB-Sr and other methods, and the distribution of the predicted values is more focused, which proves the effectiveness of the method proposed in this paper.

However, BRB-SAQF, as a derived method of BRB, has the following disadvantages. First, when the method has more premise attributes or reference points for the attributes, BRB-SAQF causes the problem of rule combination explosion, i.e., the number of rules is huge. As a result, initializing the parameters is more complicated. Therefore, in our future research, we will focus on the following aspects:

- 1) Research and design methods to reduce the number of rules to solve the problem of rule combination explosion.
- 2) Design the belief rule base for the network structure so that the number of rules per submodule can be reduced while reducing the difficulty of initializing the parameters.
- 3) Research and design the construction method of a network structure BRB so that the structure of the model can be more consistent with the working mechanism of the diagnosed object.

Funding Statement: This work was partly supported by the Postdoctoral Science Foundation of China under Grant No. 2020M683736, partly by the Teaching reform project of higher education in Heilongjiang Province under Grant No. SJGY20210456, partly by the Natural Science Foundation of Heilongjiang Province of China under Grant No. LH2021F038, partly by the Haiyan foundation of

Harbin Medical University Cancer Hospital under Grant No. JJMS2021-28, and partly by the graduate academic innovation project of Harbin Normal University under Grant Nos. HSDSSCX2022-17, HSDSSCX2022-18 and HSDSSCX2022-19. The recipient of all the above funds is W. He.

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

References

- [1] M. Angurala and V. Khullar, "A survey on various congestion control techniques in wireless sensor networks," *International Journal on Recent and Innovation Trends in Computing and Communication*, vol. 10, no. 8, pp. 47–54, 2022.
- [2] P. Patil, Waghole, Deshpande and Karykarte, "Sectoring method for improving various QoS parameters of wireless sensor networks to improve lifespan of the network," *International Journal on Recent and Innovation Trends in Computing and Communication*, vol. 10, no. 6, pp. 37–43, 2022.
- [3] M. S. Rajan, G. Dilip, N. Kannan, M. Namratha, S. Majji *et al.*, "Diagnosis of fault node in wireless sensor networks using adaptive neuro-fuzzy inference system," *Applied Nanoscience*, 2021. <https://doi.org/10.1007/s13204-021-01934-0>.
- [4] M. Zhao, Z. Tian and T. W. S. Chow, "Fault diagnosis on wireless sensor network using the neighborhood kernel density estimation," *Neural Computing and Applications*, vol. 31, no. 8, pp. 4019–4030, 2019.
- [5] U. Saeed, S. U. Jan, Y. D. Lee and I. Koo, "Fault diagnosis based on extremely randomized trees in wireless sensor networks," *Reliability Engineering & System Safety*, 2021. <https://doi.org/10.1016/j.res.2020.107284>.
- [6] Z. Noshad, N. Javaid, T. Saba, Z. Wadud, M. Q. Saleem *et al.*, "Fault detection in wireless sensor networks through the random forest classifier," *Sensors*, vol. 19, no. 7, pp. 1568, 2019.
- [7] E. Moridi, M. Haghparast and M. Hosseinzadeh, "Novel fault management framework using markov chain in wireless sensor networks: FMMC," *Wireless Personal Communications*, vol. 114, no. 1, pp. 583–608, 2020.
- [8] R. R. Swain, M. K. Pabitra and D. Tirtharaj, "Multifault diagnosis in WSN using a hybrid metaheuristic trained neural network," *Digital Communications and Networks*, vol. 6, no. 1, pp. 86–100, 2018.
- [9] M. M. Gharamaleki and B. Shahram, "A new distributed fault detection method for wireless sensor networks," *IEEE Systems Journal*, vol. 14, no. 4, pp. 4883–4890, 2020.
- [10] S. Mohapatra, P. M. Khilar and R. R. Swain, "Fault diagnosis in wireless sensor network using clonal selection principle and probabilistic neural network approach," *International Journal of Communication Systems*, vol. 32, no. 16, 2019. <https://doi.org/10.1002/dac.4138>.
- [11] R. Regin, S. S. Rajest and B. Singh, "Fault detection in wireless sensor network based on deep learning algorithms," *EAI Transactions on Scalable Information Systems*, 2021. <https://eudl.eu/doi/10.4108/eai.3-5-2021.169578>.
- [12] R. R. Swain, P. M. Khilar and T. Dash, "Neural network based automated detection of link failures in wireless sensor networks and extension to a study on the detection of disjoint nodes," *Journal of Ambient Intelligence and Humanized Computing*, vol. 10, no. 2, pp. 593–610, 2018.
- [13] A. Javaid, N. Javaid, Z. Wadud, T. Sava, O. E. Sheta *et al.*, "Machine learning algorithms and fault detection for improved belief function based decision fusion in wireless sensor networks," *Sensors*, vol. 19, no. 6, pp. 1334, 2019.
- [14] 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 Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans*, vol. 36, no. 2, pp. 266–285, 2006.
- [15] Z. Zhou, Y. Cao, G. Hu, Y. Zhang, S. Tang *et al.*, "New health-state assessment model based on belief rule base with interpretability," *Science China Information Sciences*, vol. 64, no. 7, pp. 1–15, 2021.
- [16] Z. J. Zhou, G. Y. Hu, C. H. Hu, C. L. Wen and L. L. Chang, "A survey of belief rule-base expert system," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 51, no. 8, pp. 4944–4958, 2019.

- [17] G. Wang, Y. Cui, J. Wang, L. Wu and G. Hu, "A novel method for detecting advanced persistent threat attack based on belief rule base," *Applied Sciences*, vol. 11, no. 21, pp. 9899, 2021.
- [18] Z. Feng, Z. Zhou, C. Hu, X. Ban and G. Hu, "A safety assessment model based on belief rule base with new optimization method," *Reliability Engineering & System Safety*, vol. 203, 2020. <https://doi.org/10.1016/j.res.2020.107055>.
- [19] H. L. Zhu, S. S. Liu, Y. Y. Qu, X. X. Han, W. He *et al.*, "A new risk assessment method based on belief rule base and fault tree analysis," *Proceedings of the Institution of Mechanical Engineers, Part O: Journal of Risk and Reliability*, vol. 236, no. 3, pp. 420–438, 2021.
- [20] Y. Cao, Z. J. Zhou, C. H. Hu, S. W. Tang and J. Wang, "A new approximate belief rule base expert system for complex system modelling," *Decision Support Systems*, vol. 150, 2021. <https://doi.org/10.1016/j.dss.2021.113558>.
- [21] W. He, P. L. Qiao, Z. J. Zhou, G. Y. Hu, Z. C. Feng *et al.*, "A new belief-rule-based method for fault diagnosis of wireless sensor network," *IEEE Access*, vol. 6, pp. 9404–9419, 2018.
- [22] Z. Feng, Z. J. Zhou, C. Hu, L. L. Chang, G. Y. Hu *et al.*, "A new belief rule base model with attribute reliability," *IEEE Transactions on Fuzzy Systems*, vol. 27, no. 5, pp. 903–916, 2018.
- [23] D. N. Joanes and C. A. Gill, "Comparing measures of sample skewness and kurtosis," *Journal of the Royal Statistical Society (Series D): The Statistician*, vol. 47, no. 1, pp. 183–189, 1998.
- [24] W. He, L. C. Liu and J. P. Yang, "Reliability analysis of stiffened tank-roof stability with multiple random variables using minimum distance and lagrange methods," *Engineering Failure Analysis*, vol. 32, pp. 304–311, 2013.
- [25] B. K. Lad and M. S. Kulkarni, "A parameter estimation method for machine tool reliability analysis using expert judgement," *International Journal of Data Analysis Techniques and Strategies*, vol. 2, no. 2, pp. 155–169, 2010.
- [26] X. B. Xu, J. Zheng, D. L. Xu and J. B. Yang, "Information fusion method for fault diagnosis based on evidential reasoning rule," *Journal of Control Theory and Applications*, vol. 32, pp. 1170–1182, 2015.
- [27] Z. J. Zhou, C. H. Hu, G. Y. Hu, X. X. Han, B. C. Zhang *et al.*, "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.
- [28] Z. Feng, W. He, Z. Zhou, X. Ban, C. Hu *et al.*, "A new safety assessment method based on belief rule base with attribute reliability," *IEEE/CAA Journal of Automatica Sinica*, vol. 8, no. 11, pp. 1774–1785, 2020.
- [29] X. Cheng, S. Liu, W. He, P. Zhang, B. Xu *et al.*, "A model for flywheel fault diagnosis based on fuzzy fault tree analysis and belief rule base," *Machines*, vol. 10, no. 2, pp. 73, 2022.
- [30] H. Zhu, W. Geng and J. Hanm, "Constructing a WSN node fault detection model using the belief rule base," *CAAI Transactions on Intelligent Systems*, vol. 16, no. 3, pp. 511–517, 2021.
- [31] H. Zhu, J. Li, Y. Gao and G. Cheng, "Consensus analysis of UAV swarm cooperative situation awareness," in *Proc. ICVRIS*, Zhangjiajie, Hunan, China, pp. 415–418, 2020.
- [32] Y. C. Ho, "On the perturbation analysis of discrete-event dynamic systems," *Journal of Optimization, Theory and Applications*, vol. 46, no. 4, pp. 535–545, 1985.
- [33] N. Ramanathan, E. Kohler and D. Estrin, "Towards a debugging system for sensor networks," *International Journal of Network Management*, vol. 15, no. 4, pp. 223–234, 2005.