

Article

Fault Diagnosis for Complex Equipment Based on Belief Rule Base with Adaptive Nonlinear Membership Function

Zheng Lian ^{1,*}, Zhijie Zhou ¹, Xin Zhang ¹, Zhichao Feng ¹, Xiaoxia Han ² and Changhua Hu ¹¹ Missile Engineering Institute, PLA Rocket Force University of Engineering, Xi'an 710025, China² College of War Support, PLA Rocket Force University of Engineering, Xi'an 710025, China

* Correspondence: lz09187910@163.com

Abstract: Fault diagnosis of complex equipment has become a hot field in recent years. Due to excellent uncertainty processing capability and small sample problem modeling capability, belief rule base (BRB) has been widely used in the fault diagnosis. However, previous BRB models almost did not consider the diverse distributions of observation data which may reduce diagnostic accuracy. In this paper, a new fault diagnosis model based on BRB is proposed. Considering that the previous triangular membership function cannot address the diverse distribution of observation data, a new nonlinear membership function is proposed to transform the input information. Then, since the model parameters initially determined by experts are inaccurate, a new parameter optimization model with the parameters of the nonlinear membership function is proposed and driven by the gradient descent method to prevent the expert knowledge from being destroyed. A fault diagnosis case of laser gyro is used to verify the validity of the proposed model. In the case study, the diagnosis accuracy of the new BRB-based fault diagnosis model reached 95.56%, which shows better fault diagnosis performance than other methods.

Keywords: fault diagnosis; belief rule base; membership function; model optimization



Citation: Lian, Z.; Zhou, Z.; Zhang, X.; Feng, Z.; Han, X.; Hu, C. Fault Diagnosis for Complex Equipment Based on Belief Rule Base with Adaptive Nonlinear Membership Function. *Entropy* **2023**, *25*, 442. <https://doi.org/10.3390/e25030442>

Academic Editor: Katalin M. Hangos

Received: 2 February 2023

Revised: 23 February 2023

Accepted: 25 February 2023

Published: 2 March 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

With the rapid development of industrial technology, the fault diagnosis of complex equipment has received extensive attention and become a hot topic in Prognostics Health Management (PHM) [1–3]. At present, due to the complexity and intelligence of complex equipment in industrial production and national defense science and industry, it is tough to diagnose their faults through the appearance of the equipment. Therefore, establishing a fault diagnosis model is an effective method.

Current popular fault diagnosis models can be divided into three categories: mechanism-based models, knowledge-based models and data-driven models [4]. (1) Mechanism-based model: such models require that the mechanism of complex equipment can be clearly understood and corresponding mathematical or physical models can be established, such as the Kalman filter [5] and Kirchhoff law [6]. However, with the increasing complexity of equipment, its mechanism is difficult to grasp and such models are rarely used. (2) Knowledge-based model: this kind of model is established by the qualitative knowledge of domain experts, but it is generally difficult to achieve high modeling accuracy such as fault tree analysis (FTA) [7] and analytic hierarchy process (AHP) [8]. Since such models cannot learn by themselves, when the equipment is running, it cannot update knowledge and can only be modified by experts. (3) Data-driven model: with the advent of the big data era, data-driven fault diagnosis models have attracted a lot of attention. This kind of model does not need to master the mechanics of the equipment and can achieve modeling of the input and output relationship through the observation data. Simultaneously, they also have good learning abilities and can constantly be updated by data, such as with the support vector machines [9], decision trees [10] and deep learning algorithms [11]. However, in many fault diagnosis fields, the high-value fault

samples are very limited, causing the data-driven model to fall into the problem of overfitting. On the other hand, due to the black-box nature of most data-driven models, the diagnosis results are not transparent and difficult to be convincing.

Belief rule base (BRB) is a generic rule-based modeling approach that was formally proposed by Yang et al. in 2006 [12]. It differs from traditional ML-type fuzzy rules: the consequent part of the “if-then” rule in BRB is composed of the belief distribution of all possible results, which means that BRB can handle fuzzy, uncertain and incomplete information together. Since rule-based knowledge representation is easy to understand by experts, BRB can embed expert knowledge easily and has good interpretability [13]. Therefore, as a knowledge-data hybrid-driven modeling method [14], BRB has been widely used in the field of fault diagnosis [15]. Xu et al. developed a fault diagnosis model of diesel engines based on BRB for the first time and realized the diagnosis of concurrent faults [16]. Feng et al. established a fault diagnosis model for oil pipelines based on BRB after considering the correlation of attributes [17]. Zhang et al. established the fault diagnosis and location model of the bus network based on BRB [18]. Li et al. developed an adaptive interpretable fault diagnosis model based on BRB [19]. Chen et al. proposed a mechanical equipment fault diagnosis model based on combination BRB [20]. Ming et al. proposed an interpretable fault diagnosis method based on probability table and BRB [21]. Wang et al. proposed a BRB-based fault diagnosis method for multi-agent systems [22].

As a typical fuzzy system, the first step of BRB in fault diagnosis is the transformation of input information. Since the observation information of most fault indicators in engineering are quantitative data, it needs to be fuzzified. In the previous BRB-based fault diagnosis model, the rule and utility based method is used to fuzzy quantitative data [23]: this is a triangular membership function [24]. The triangular membership function has a simple structure and can accurately transform the uniformly distributed observation data. However, in engineering, many observation data are not subject to the uniform distribution. Moreover, the small sample characteristics of fault data may present a distribution form that is difficult to be directly understood by people. In these cases, the triangular membership function is not accurate enough and will reduce the diagnostic accuracy. Therefore, it is necessary to adopt a membership function that can adapt the data distribution to transform the input information to improve the modeling accuracy.

As a commonly used membership function, the gaussian membership function can adapt to the distribution of data by adjusting its expectation and variance. Some research on this membership function in BRB has been carried out. Liu et al. introduced the Gaussian membership function in BRB-based engineering system safety analysis [25]; Zhang et al. also considered the Gaussian membership function when establishing the fault diagnosis model of the bus network [18]. However, it is worth noting that the adaptive ability of the Gaussian membership function is limited by the shape of the exponential curve. The Gaussian membership function is not ideal for the transformation of uniformly distributed observation data.

Therefore, a nonlinear membership function is proposed in this paper to make up for the shortcomings of the Gaussian membership function. In the exponential part of the original triangular membership function, a new parameter is considered to control the shape of the membership function curve. When the parameters change, the nonlinear membership function can adaptively transform the observation data under various distributions. Since the shape of the curve of the function changes more flexibly, the nonlinear membership function can more accurately convert the input information than the Gaussian membership function.

On the other hand, as an expert system, parameters of BRB model are initially determined by experts. Due to the inherent subjectivity and ignorance of experts, the diagnostic accuracy of the initial BRB is often not ideal and needs to be optimized. In recent years, many swarm intelligent algorithms have been developed for BRB parameter optimization, such as DE [26], PSO [27] and P-CMAES [28]. The common problem of these swarm intelligence algorithms is that expert knowledge is destroyed after optimization due to the

random initialization of the population. This will cause BRB to lose its advantage in fault diagnosis. However, the gradient descent method searches from the parameters initially determined by experts, which allows retention of expert knowledge to the greatest extent. Therefore, a new optimization model based on the gradient descent method, which can further improve the modeling accuracy while maintaining expert knowledge, is developed. Therefore, the contributions of this paper are the following:

- (1) A new BRB considering nonlinear membership function, which can adaptively deal with the non-uniform distribution of observation data in fault diagnosis, is proposed.
- (2) For the new BRB model, a new optimization model based on the gradient descent method is proposed to improve the accuracy of fault diagnosis and keep expert knowledge from being destroyed.

The structure of this paper is as follows: In Section 2, the problem and basic knowledge are introduced. In Section 3, a new BRB-based fault diagnosis model with an adaptive membership function is proposed. In Section 4, a fault diagnosis case of the laser gyro is used to verify the validity of the proposed model. This paper is summarized in Section 5.

2. Problem Description and Basic Knowledge

2.1. Problem Description

Faced with the characteristics of few high-value fault samples and existing expert knowledge in the fault diagnosis for the complex equipment, this paper mainly solves the following three problems:

Problem 1: How to use a small amount of test data and existing expert knowledge to establish the fault diagnosis model. Due to the working characteristics of most complex equipment, the capacity of its observation data is very limited. Therefore, the data-driven fault diagnosis model is prone to overfitting since it cannot reflect all fault modes. On the other hand, in the long-term operation of the equipment, experts in the field have accumulated abundant expert knowledge that can be used to help judge the fault mode of the complex equipment. Therefore, the following mapping relationships need to be established:

$$Y = R(x_1, x_2, \dots, x_M, EK) \tag{1}$$

where $[x_1, x_2, \dots, x_M]$ represent M fault indicators. Y is the fault mode to be diagnosed. $R(\cdot)$ is the mapping function of the fault diagnosis model. EK is the expert knowledge.

Problem 2: For Problem 1, a belief rule-based fault diagnosis model will be proposed in Section 2.2. In engineering, observation data usually do not obey the uniform distribution. However, the triangular membership function in the previous BRB-based fault diagnosis model cannot accurately convert the non-uniformly distributed data, which will lead to poor fault diagnosis accuracy. Therefore, a new fuzzy membership function is needed to reflect the influence of uneven data on information transformation.

Problem 3: As an expert system, the parameters in the membership function of BRB are generally determined intuitively by experts according to domain knowledge and the distribution characteristics of observation data. However, due to the subjectivity and fuzziness of expert knowledge, these parameters cannot be given accurately, leading to a decrease in the modeling accuracy. Therefore, it is necessary to further improve the accuracy of fault diagnosis through parameter optimization.

2.2. Belief Rule-Based Fault Diagnosis Model

As a generalized ML-type fuzzy system, BRB is composed of a series of “if-then” belief rules, as shown below:

$$R_k : \text{ IF } x_1 \text{ is } H_1^k \wedge x_2 \text{ is } H_2^k \wedge \dots \wedge x_M \text{ is } H_M^k, \\ \text{ THEN } \{(D_1, \beta_{1,k}), (D_2, \beta_{2,k}) \dots (D_N, \beta_{N,k})\}, (\sum_{n=1}^N \beta_{n,k} \leq 1), \\ \text{ with a rule weight } \theta_k (k = 1, 2 \dots, L) \text{ and attribute weight } \delta_i (i = 1, 2 \dots, T_k) \tag{2}$$

where $[x_1, x_2, \dots, x_M]$ is the input vector and consists of m dimensions components. In fault diagnosis, x_i is the i th fault indicator. $H_i^k (i = 1, 2, \dots, T_k)$ is the referential value of the $U_i (i = 1, 2, \dots, T_k)$ attribute in the k th rule. It is generally determined by experts according to industry standards or observation data. L is the number of rules. $D_n (n = 1, 2, \dots, N)$ represents N possible faults. $\beta_{n,k} (n = 1, 2, \dots, N)$ represents the belief degree of the n th fault D_n , reflecting the support of the k th rule for this consequence. θ_k is the rule weight of the k th rule, which represents the relative importance of each rule. δ_i is the weight of the i th attribute, which reflects the relative importance of fault indicators.

Benefited from the knowledge representation based on the belief rule, BRB has the following advantages in fault diagnosis:

(1) Expert knowledge. Due to the natural advantages of language models, BRB uses rules to represent the nonlinear mapping relationship between fault indicators and fault modes, enabling experts and users to easily understand the behavior of the model. Therefore, compared with neural networks, support vector machines and decision tree models, the initial parameters of BRB can be determined by experts based on domain knowledge and existing observation data and the model is able to roughly reflect the mapping relationship of the system.

(2) Small sample modeling ability. The fault data of much equipment are characterized by high values and a small number of samples. Fortunately, due to the embedding of expert knowledge, BRB can model the system comprehensively with very limited observation data, even if the initial mapping relationship is rough. This means that BRB will not fall into the overfitting problem like the data-driven model. On the other hand, Chen et al. proved that BRB is a general approximation model [29], so this model has ideal modeling accuracy.

(3) Ability to process uncertain information. Compared with general fuzzy systems, BRB extends the “then” part of the rule to the belief distribution of all possible results, enabling BRB to deal with the probability uncertainty while solving the fuzzy uncertainty. Therefore, BRB can also show good performance when partial observation data are missing in the engineering application.

3. The Proposed Method

In this section, a BRB model with an adaptive nonlinear membership function is proposed to model the fault diagnosis considering non-uniform distribution observation data. In Section 3.1, the shortcomings of the existing triangular membership function and Gaussian membership function are analyzed in detail. Then, a nonlinear membership function is proposed to solve Problem 2. In Section 3.2, based on the gradient descent method, an optimization model considering the parameters of the nonlinear membership function is proposed to solve Problem 3.

3.1. Inference Process Based on the Nonlinear Membership Function

In fault diagnosis, all belief rules constitute a knowledge base. After the input information of the fault indicator is obtained, the fault mode can be diagnosed based on this knowledge base. It is worth noting that, as a fuzzy system, belief rules in BRB are expressed as mappings to linguistic values. However, the observation data of fault indicators are mostly quantitative information. Therefore, it is necessary to convert quantitative observation data into membership degree of all linguistic referential grades through a membership function, which is so-called “fuzzification” as follows:

$$S(x_i) = \{(H_{ij}, \alpha_{ij}); j = 1, \dots, J_i\}, \quad i = 1, 2, \dots, M \quad (3)$$

where H_{ij} is the j th referential grade of the i th fault indicator and α_{ij} is the corresponding membership degree. x_i are the quantitative observation data.

In previous BRB models, the most commonly applied membership function is the triangular membership function, which is used in rule (or utility) based transformation methods, as shown below:

$$\alpha_{ij} = \begin{cases} \frac{a_i^{(k+1)} - x_i}{a_i^{(k+1)} - a_{ik}}, & j = k \text{ if } a_{ik} \leq A_i^* \leq a_i^{(k+1)} \\ \frac{x_i - a_{ik}}{a_i^{(k+1)} - a_{ik}}, & j = k + 1 \\ 0, & j = 1, 2, \dots, L, j \neq k, k + 1 \end{cases} \quad (4)$$

where α_{ij} is a quantitative value corresponding to H_{ij} , which is usually determined by experts.

The curve of the above membership function is shown in Figure 1. It can be seen from Figure 1 and Equation (4) that, since the derivative of this function is constant, the changing trend of membership of each referential grade is a straight line. However, when the observation data are uneven, such as when the data are concentrated in a certain area, this membership function cannot accurately reflect the corresponding membership, as shown in Figure 2. It can be seen from Figure 2a that the observation data are distributed evenly between the two referential grades. Therefore, it is easy to understand that the change in membership degree is linear in this case. But, in Figure 2b, the observation data are concentrated near the referential grade H_{n+1} . Therefore, for the two points marked by the red dotted line, their membership degree distribution should be different. The membership of the yellow point is assigned as $H_n = H_{n+1} = 0.5$. For the green point in Figure 2b, since this point is closer to H_n in the whole dataset, the membership degree of H_n shall be greater than 0.5 and the membership degree of H_{n+1} shall be less than 0.5, correspondingly.

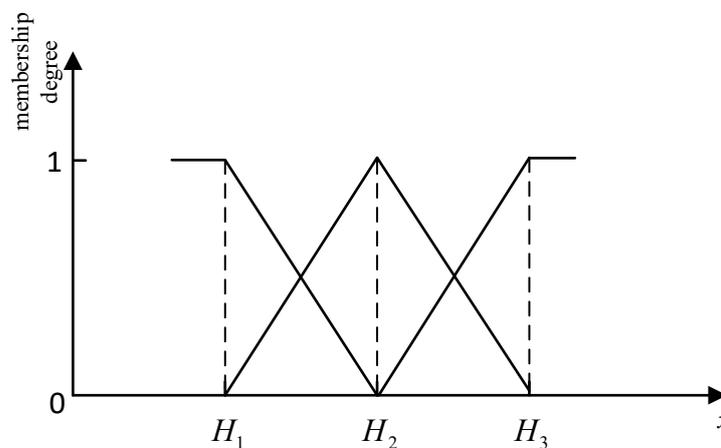


Figure 1. Membership degree using the triangular membership function.

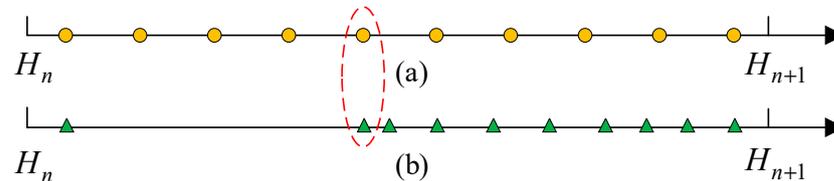


Figure 2. Comparison of uneven data distribution. The uniformly distributed data is shown in (a), and the non-uniformly distributed data is shown in (b).

For example, the two referential grades in Figure 2 correspond to the semantic values “low” and “high”, respectively. For the point marked in red in Figure 2b, it should be a lower value in the entire dataset. Therefore, the membership degree of the referential grade “low” should be higher.

The inaccurate quantitative data fuzzification will reduce the modeling accuracy of the fault diagnosis model. Thus, a nonlinear membership function is proposed for the fuzzification of uneven quantitative data in this paper, which can be described as follows:

$$\alpha_{ij} = \begin{cases} \left(\frac{a_{i(k+1)}-x_i}{a_{i(k+1)}-a_{ik}}\right)^s, & j = k \text{ if } a_{ik} \leq A_i^* \leq a_{i(k+1)} \\ 1 - \left(\frac{a_{i(k+1)}-x_i}{a_{i(k+1)}-a_{ik}}\right)^s, & j = k + 1 \\ 0, & j = 1, 2, \dots, L, j \neq k, k + 1 \end{cases} \quad (5)$$

where $s \in (0, +\infty)$ is the parameter of the function, which can reflect the distribution of observation data.

With the change of s , the new membership function can adaptively reflect the impact of different distributions of data. For example, when s is 0.25, 0.5, 1, 2 and 5, respectively, the curves of the membership function are shown in Figure 3. It can be seen that with the increase of n , the function changes from convex to concave. In particular, when s equals 1, the nonlinear membership function degenerates into a triangular membership function. Correspondingly, for the distribution of data in Figure 2b, the nonlinear membership function at $s = 0.25$ or 0.5 can more accurately conduct the fuzzification of quantitative data in this case.

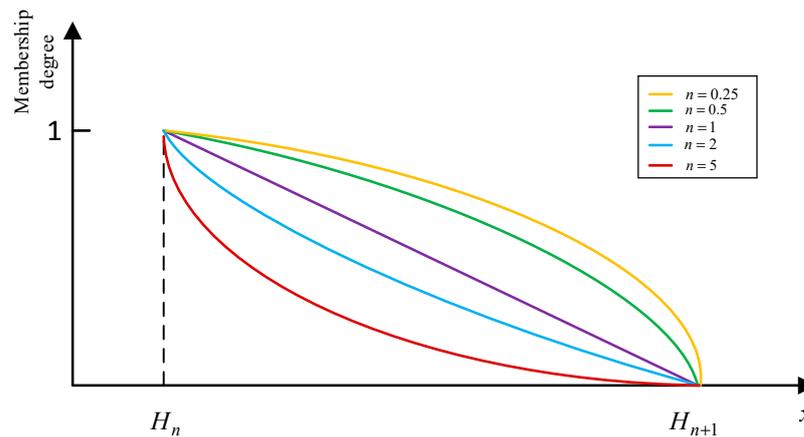


Figure 3. Membership degree of H_n using the nonlinear membership function.

It is worth noting that, as another commonly used membership function in fuzzy systems, the Gaussian membership function can also realize adaptive fuzzification of input data through changes in expectation and standard deviation, as shown below:

$$\alpha_{ij} = \exp\left(-\frac{1}{2}\left(\frac{x_i - c_{ij}}{\sigma_{ij}}\right)^2\right) \quad (6)$$

where c_{ij} is the expectation and σ_{ij} is the standard deviation.

However, it has the following two shortcomings: firstly, the Gaussian membership function cannot achieve accurate transformation of uniformly distributed input information, which is limited by the characteristics of its nonlinear curve. However, when $s = 1$, the nonlinear membership function proposed in this paper can avoid this problem. Secondly, the adaptive ability of the Gaussian membership function is insufficient. For the data distribution within an interval, the Gaussian membership function can only self-adapt the data distribution under partial circumstances, as shown in Figure 4. The standard variance of the Gaussian membership function is 0.25, 0.5, 1, 2 and 5, respectively, in Figure 4. With the increase of variance, the shape of the exponential function curve cannot properly reflect the distribution characteristics of the dataset close to H_{n+1} . Furthermore, when the input is at H_{n+1} , the membership degree of H_n is still quite high, which is difficult for users

to understand. Therefore, the Gaussian membership function is only applicable when the observation data are concentrated near the referential grade H_n . Based on the above analysis, it can be seen that the nonlinear membership function proposed in this paper can more accurately reflect the distributions of data.

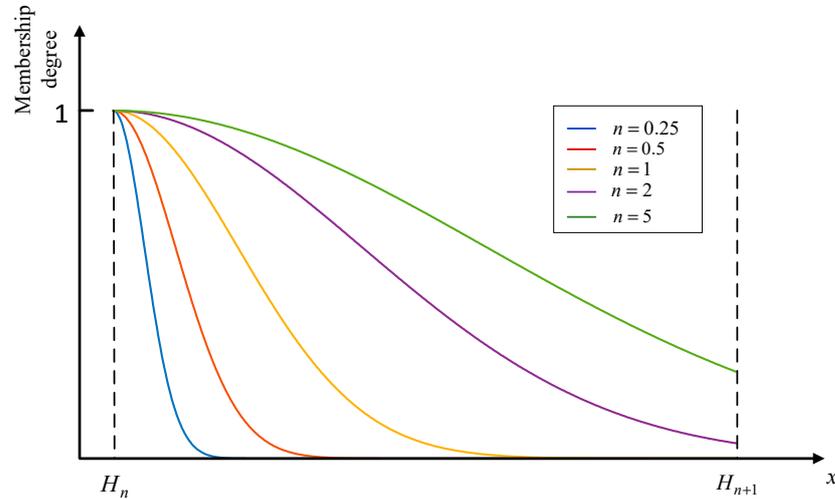


Figure 4. Membership degree of H_n using the Gaussian membership function.

Therefore, based on the nonlinear membership function, when observation data are obtained, the steps for fault diagnosis can be described as follows:

Step 1: Fuzzification of quantitative data. The referential grade of each fault indicator is a fuzzy partition, which is assigned to the nonlinear membership function $R(\cdot)_{ij}$. For the observation data of i th fault indicator, the membership degree of each referential grade is calculated as follows:

$$\alpha_{ij} = R_{ij}(x_i) = \begin{cases} \left(\frac{a_{i(k+1)} - x_i}{a_{i(k+1)} - a_{ik}}\right)^{s_{ij}}, & j = k \text{ if } a_{ik} \leq x_i \leq a_{i(k+1)} \\ 1 - \left(\frac{a_{i(k+1)} - x_i}{a_{i(k+1)} - a_{ik}}\right)^{s_{ij}}, & j = k + 1 \\ 0, & j = 1, 2, \dots, L, j \neq k, k + 1 \end{cases} \quad (7)$$

where x_i are the input data. s_{ij} is the parameter of the nonlinear membership function, which is usually determined by experts after observing the distribution of data or calculated based on statistical methods.

Step 2: Activation of belief rules. The activation weight of the rule is calculated as follows:

$$w_k = \frac{\theta_k \alpha_k}{\sum_{l=1}^L \theta_l \alpha_l} \quad (8)$$

where

$$\alpha_k = \prod_{i=1}^{M_k} (\alpha_i^k)^{\bar{\delta}_i}, \bar{\delta}_i = \frac{\delta_i}{\max_{i=1,2,\dots,M_k} \{\delta_i\}} \quad (9)$$

Step 3: Reasoning of activated rules. In this paper, the analytic ER algorithm [30] is used to fuse the activated rules to obtain the belief degree of each failure mode as follows:

$$\hat{\beta}_n = \frac{\mu[\prod_{k=1}^L (w_k \beta_{n,k} + 1 - w_k \sum_{j=1}^N \beta_{j,k}) - \prod_{k=1}^L (1 - w_k \sum_{j=1}^N \beta_{j,k})]}{1 - \mu[\prod_{k=1}^L (1 - w_k)]}, \quad (10)$$

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

$$\mu = [\sum_{n=1}^N \prod_{k=1}^L (w_k \beta_{n,k} + 1 - w_k \sum_{j=1}^N \beta_{j,k}) - (N - 1) \prod_{k=1}^L (1 - w_k \sum_{j=1}^N \beta_{j,k})]^{-1}$$

where $\hat{\beta}_n$ represents the belief degree of the n th failure mode D_n .

In general, the failure mode with the highest belief degree is regarded as a possible failure as the output of the model as follows:

$$\hat{n} = \operatorname{argmax}_n(\hat{\beta}_n) \tag{11}$$

where \hat{n} indicates the diagnosed fault mode.

3.2. Model Optimization Based on the Gradient Descent Method

Due to the subjectivity and fuzziness of expert knowledge, the modeling accuracy of the initially constructed fault diagnosis model is generally difficult to meet the requirements of practical engineering. Therefore, the model parameters initially determined by experts in the BRB need to be optimized to improve the diagnostic accuracy of the model. In general, for classification problems such as fault diagnosis, the cross-entropy loss function is used as the objective function as follows:

$$Q(\Omega) = -\frac{1}{T} \sum_{t=1}^T \sum_{j=1}^N y_j^t \log_2 \hat{\beta}_j^t, y_j^t = \begin{cases} 0(j \neq \hat{y}_t) \\ 1(j = \hat{y}_t) \end{cases} \tag{12}$$

where $\hat{y} \in \{1, 2, \dots, N\}$ indicates the category of the real fault. T is the capacity of observation data. $\Omega = \{\theta, \delta, \beta, s\}$ is a parameter vector, which is composed of rule weight, attribute weight, basic belief degree and parameters of the membership function.

Considering the constraint conditions of parameters in BRB model, the following parameter optimization model can be constructed:

$$\begin{aligned} \min Q(\{\theta, \delta, \beta, s\}) &= -\frac{1}{T} \sum_{t=1}^T \sum_{j=1}^N y_j^t \log_2 \hat{\beta}_j^t, y_j^t = \begin{cases} 0(j \neq \hat{y}_t) \\ 1(j = \hat{y}_t) \end{cases} \\ \text{s.t.} & \\ 0 \leq \theta_k \leq 1, 0 \leq \delta_i \leq 1, 0 \leq \beta_{n,k} \leq 1, \sum_{n=1}^N \beta_{n,k} \leq 1, s_{ij} > 0 & \\ (k = 1, 2, \dots, L, i = 1, 2, \dots, M, n = 1, 2, \dots, N, j = 1, 2, \dots, J_i) & \end{aligned} \tag{13}$$

In recent years, many optimization algorithms have been developed for BRB model parameter optimization, such as DE, PSO, P-CMAES and other swarm intelligence algorithms. Yang et al. [31] pointed out that when BRB is used as an expert system, the optimization of model parameters should only be “fine-tuning”, which is also a major difference between BRB and artificial neural networks. Feng et al. [32] pointed out that due to the operation of population initialization of swarm intelligence algorithm, the expert knowledge in BRB is likely to be destroyed and the reasoning results may conflict with intuition. This may cause the fault diagnosis results to be difficult to be convincing and weaken the interpretability of the model. Compared with the swarm intelligence algorithm, the gradient descent method directly uses derivative information and takes the parameters initially determined by experts as the initial value of optimization to search, retaining initial expert knowledge to the greatest extent. Therefore, in this paper, stemming from the derivability of the BRB reasoning process, an optimization algorithm based on gradient descent is used to train the model.

There are 4 types of parameters as optimization variables. Therefore, it is necessary to calculate the first-order partial derivative of the objective function with respect to them.

First, the first-order partial derivative of the objective function Q with respect to the reasoning result $\hat{\beta}_n$ is calculated as follows:

$$\frac{\partial Q}{\partial \hat{\beta}_n} = -\frac{1}{T} \sum_{t=1}^T y_j^t \frac{1}{\hat{\beta}_n^t}, y_j^t = \begin{cases} 0 & (j \neq \hat{y}_t) \\ 1 & (j = \hat{y}_t) \end{cases} \tag{14}$$

The first-order partial derivative of the reasoning result $\hat{\beta}_n$ with respect to the basic belief degree $\beta_{r,f}$ is:

$$\frac{\partial \hat{\beta}_n}{\partial \beta_{r,f}} = \begin{cases} \left[\frac{\left[\sum_{p=1}^N \zeta(p) - (N-1) \prod_{k=1}^L (1-w_k) \right] \zeta(n) \frac{w_f}{w_f \beta_{r,f} + 1 - w_f} \prod_{k=1}^L (1-w_k)}{\left[\sum_{p=1}^N \zeta(p) - N \right]} \right]^2, & r = n \\ \frac{\left[\prod_{k=1}^L (1-w_k) - \zeta(n) \right] \zeta(r) \frac{w_f}{w_f \beta_{r,f} + 1 - w_f}}{\left[\sum_{p=1}^N \zeta(p) - N \prod_{k=1}^L (1-w_k) \right]^2}, & r \neq n \end{cases} \quad (15)$$

where

$$\zeta(p) = \prod_{k=1}^L (w_k \beta_{p,k} + 1 - w_k) \quad (16)$$

So far, the first-order partial derivative of the first type of parameter has been calculated as follows:

$$\frac{\partial Q}{\partial \beta_{r,k}} = \sum_{n=1}^N \frac{\partial Q}{\partial \hat{\beta}_n} \frac{\partial \hat{\beta}_n}{\partial \beta_{r,k}} \quad (17)$$

Then, we need to calculate the first-order partial derivative of rule weight, attribute weight and parameters of the membership function. According to the chain rule, the first-order partial derivative of the reasoning result $\hat{\beta}_n$ with respect to the activation weight w_g needs to be obtained as follows:

$$\begin{aligned} \frac{\partial \hat{\beta}_n}{\partial w_f} &= \frac{[A(n)C(n) \sum_{p=1}^N A(p) - A(n) \sum_{p=1}^N A(p)C(p) - NBA(n)C(n) - \sum_{p=1}^N A(p) - NB]^2}{\left[\prod_{k=1}^L (1-w_k) \sum_{p=1}^N A(p) + NA(n) \prod_{k=1}^L (1-w_k) + B \sum_{p=1}^N A(p)C(p) \right]^2} \\ &= \frac{\prod_{k \neq f}^L (1-w_k) \sum_{p=1}^N A(p) + NA(n) \prod_{k \neq f}^L (1-w_k) + B \sum_{p=1}^N A(p)C(p)}{\left[\sum_{p=1}^N A(p) - NB \right]^2} \end{aligned} \quad (18)$$

where

$$\begin{aligned} A(n) &= \prod_{k=1}^L (w_k \beta_{n,k} + 1 - w_k) \\ B &= \prod_{k=1}^L (1 - w_k) \\ C(n) &= \frac{\beta_{n,t} - 1}{w_t \beta_{n,t} + 1 - w_t} \end{aligned} \quad (19)$$

The first derivative of the activation weight w_t with respect to the rule weight θ_f is calculated as follows:

$$\frac{\partial w_t}{\partial \theta_f} = \begin{cases} \frac{\sum_{l=1, l \neq t}^L \theta_l \alpha_l \alpha_f}{\left(\sum_{l=1}^L \theta_l \alpha_l \right)^2}, & t = f \\ -\frac{\theta_t \alpha_t \alpha_f}{\left(\sum_{l=1}^L \theta_l \alpha_l \right)^2}, & t \neq f \end{cases} \quad (20)$$

For the attribute weight δ_i , the normalization of this parameter in Equation (9) is nondifferentiable. Therefore, only the first order partial derivative of the normalized

attribute weight $\bar{\delta}_i$ can be calculated here. First, the first derivative of the activation weight w_t with respect to the rule matching degree α_f needs to be calculated:

$$\frac{\partial w_t}{\partial \alpha_f} = \begin{cases} \frac{\sum_{l=1, l \neq t}^L \theta_l \alpha_l \theta_f}{(\sum_{l=1}^L \theta_l \alpha_l)^2}, & t = f \\ -\frac{\theta_t \alpha_t \theta_f}{(\sum_{l=1}^L \theta_l \alpha_l)^2}, & t \neq f \end{cases} \tag{21}$$

Then, the first derivative of rule matching degree α_f with respect to normalized attribute weight $\bar{\delta}_i$ is calculated as follows:

$$\frac{\partial \alpha_f}{\partial \bar{\delta}_i} = \alpha_f \log(\alpha_i^k) \tag{22}$$

Therefore, according to the chain rule, the partial derivative of the objective function with respect to the rule weight and the normalized attribute weight can be calculated as follows:

$$\frac{\partial Q}{\partial \theta_k} = \sum_{t=1}^L \sum_{n=1}^N \frac{\partial Q}{\partial \hat{\beta}_n} \frac{\partial \hat{\beta}_n}{\partial w_t} \frac{\partial w_t}{\partial \theta_k} \tag{23}$$

$$\frac{\partial Q}{\partial \bar{\delta}_i} = \sum_{k=1}^L \sum_{t=1}^L \sum_{n=1}^N \frac{\partial Q}{\partial \hat{\beta}_n} \frac{\partial \hat{\beta}_n}{\partial w_t} \frac{\partial w_t}{\partial \alpha_k} \frac{\partial \alpha_k}{\partial \bar{\delta}_i} \tag{24}$$

Finally, the first order partial derivative of the individual membership degree α_i^k with respect to the parameters of the membership degree function s_{ij} is calculated as follows:

$$\frac{\partial \alpha_{ij}}{\partial s_{ij}} = \begin{cases} \left(\frac{a_{i(k+1)} - x_i}{a_{i(k+1)} - a_{ik}} \right)^{s_{ij}} \log\left(\frac{a_{i(k+1)} - x_i}{a_{i(k+1)} - a_{ik}} \right), & j = k \text{ if } a_{ik} \leq x_i \leq a_{i(k+1)} \\ -\left(\frac{a_{i(k+1)} - x_i}{a_{i(k+1)} - a_{ik}} \right)^{s_{ij}} \log\left(\frac{a_{i(k+1)} - x_i}{a_{i(k+1)} - a_{ik}} \right), & j = k + 1 \\ 0, & j = 1, 2, \dots, L, j \neq k, k + 1 \end{cases} \tag{25}$$

According to the chain rule, the partial derivative of the objective function with respect to the parameters of the membership function is calculated as follows:

$$\frac{\partial Q}{\partial s_{ij}} = \sum_{k=1}^L \sum_{t=1}^L \sum_{n=1}^N \frac{\partial Q}{\partial \hat{\beta}_n} \frac{\partial \hat{\beta}_n}{\partial w_t} \frac{\partial w_t}{\partial \alpha_k} \frac{\partial \alpha_k}{\partial \alpha_i^k} \frac{\partial \alpha_i^k}{\partial s_{ij}} \tag{26}$$

Therefore, the gradient vector of the optimization variable can be obtained as follows:

$$d = \left[\frac{\partial Q}{\partial \beta_{n,k}}, \frac{\partial Q}{\partial \theta_k}, \frac{\partial Q}{\partial \bar{\delta}_i}, \frac{\partial Q}{\partial s_{ij}} \right]^T \tag{27}$$

Since each parameter in the optimization model has corresponding constraints, they should be approximated to meet the optimization constraints after the parameter is updated based on the gradient. Thus, the steps of parameter optimization can be summarized as follows:

Step 1: The model parameters initially given by experts: basic belief degree, rule weight, attribute weight and parameters of membership function are taken as the initial value $z_k = z_0$.

Step 2: Calculate the gradient of the optimization variable d_k .

Step 3: The optimization variables are updated as follows:

$$z_{k+1} = z_k - \lambda d_k \tag{28}$$

where λ is the step size of iteration and is determined by the one-dimensional search method.

Step 4: Approximate projection operation. For inequality constraints of each parameter, when the value of the parameter does not meet the constraint conditions, take the adjacent bound as the approximate value. For example, if $\theta_{k+1} < 0$, then $\theta'_{k+1} = 0$. Moreover, the basic belief degree of a belief rule is normalized so that the sum is 1. Therefore, z'_{k+1} is obtained.

Step 5: Calculate the gradient vector d_{k+1} at this time. Judge whether the termination condition of the algorithm is reached. If yes, end. Otherwise, let $d_k = d_{k+1}, z_k = z'_{k+1}$ and go to Step 3.

Finally, the fault diagnosis model proposed in this paper can be summarized as shown in Figure 5.

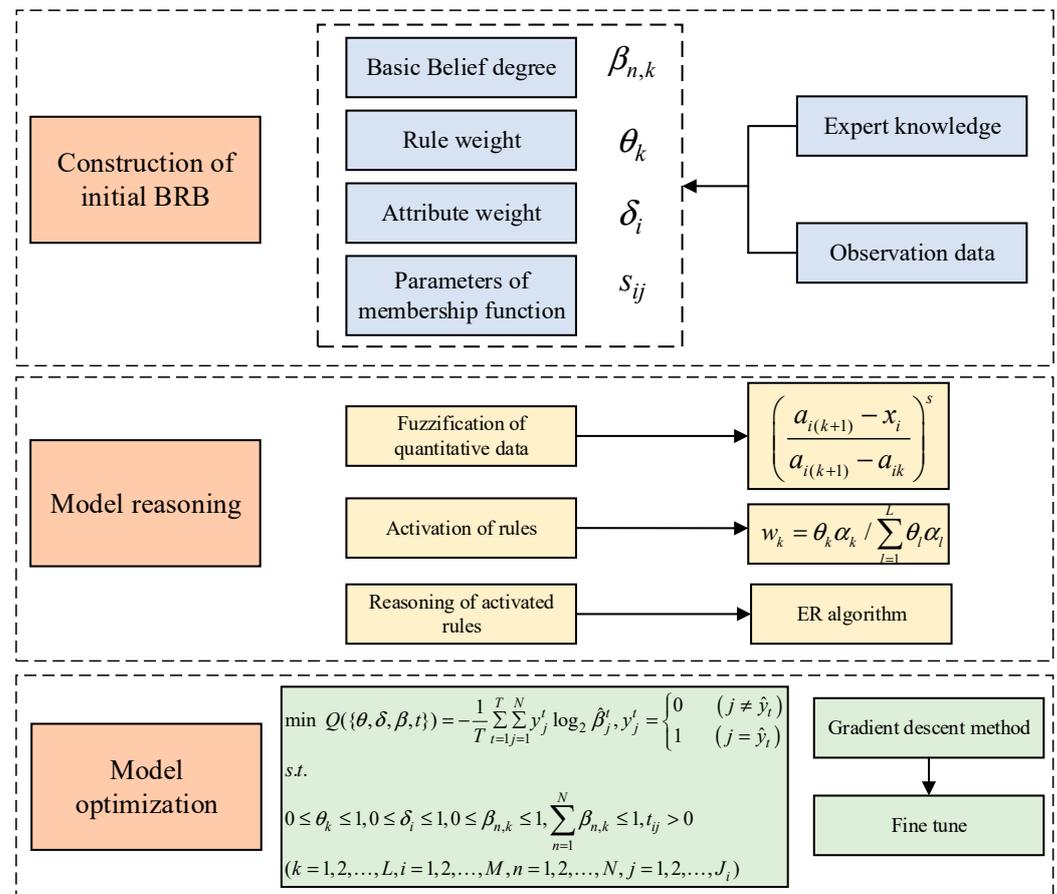


Figure 5. The whole process of the proposed model.

4. Case Study

A fault diagnosis case of a laser gyro will be used in this section to verify the effectiveness of the proposed model. In Section 4.1, the background of laser gyro fault diagnosis is briefly introduced. In Section 4.2, the BRB-based fault diagnosis model is built and optimized. In Section 4.3, a comparative study between the proposed model and other models is conducted. Analysis and discussions are carried out in Section 4.4.

4.1. Background Description

As an important navigation device, the laser gyro plays an extremely important role in many fields, such as automobiles, ships, rockets, etc. However, in the storage process of the laser gyro, due to the inevitable external interference and its own performance degradation, it is very likely to be in the fault state. Once these laser gyros are used in the failure state, it may cause unbearable personnel and property losses. For the laser gyro, it is difficult

to judge whether it is in the fault state from appearance. However, the observation data of its drift coefficient can reflect the degree of the fault. In general, the greater the drift coefficient, the higher the degree of the fault. Therefore, in this paper, the zero-order term drift coefficient D_0 , the first-order term drift coefficient D_1 and the second-order term drift coefficient D_2 are used as indicators of laser gyro fault diagnosis. For a certain laser gyro, 180 groups of observation data within a storage period are shown in Figure 6. According to the fault degree of laser gyro and the industry standard, the three fault modes are, respectively, slight fault (S), moderate fault (M) and bad fault (B). This is shown in Figure 6.

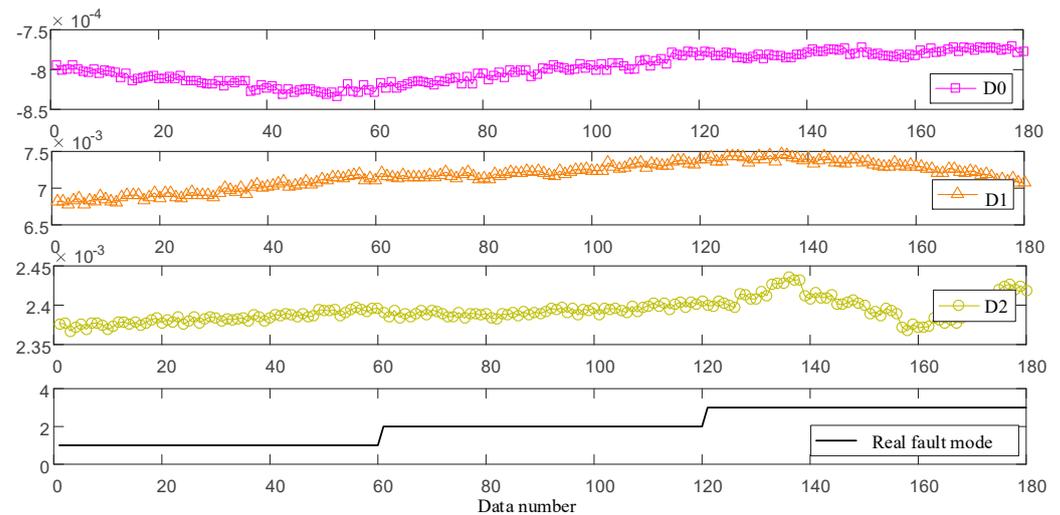


Figure 6. Observation data of fault indicators.

4.2. Construction and Optimization of the Fault Diagnosis Model

In this case study, the three drift coefficients of the laser gyro are fault indicators which are used to diagnose three types of fault modes, namely slight fault (S), moderate fault (M) and bad fault (B). Therefore, the following BRB can be established:

$$R_k: \text{ IF } D_0 \text{ is } H_1^k \wedge D_1 \text{ is } H_2^k \wedge D_2 \text{ is } H_3^k, \text{ THEN } \{ (B, \beta_{1,k}), (M, \beta_{2,k}), (S, \beta_{3,k}) \}$$

According to expert knowledge and industry standards, each fault indicator has three referential grades, namely low (L), medium (M) and high (H), whose corresponding referential values are shown in Table 1. Thus, the “then” part of the rules in the initial BRB is shown in Table 2. All rule weights are initially set to one. Since D_0 can most obviously reflect the degree of fault and D_1 takes the second place, the attribute weights are set to $\delta_1 = 1, \delta_2 = 0.7$ and $\delta_3 = 0.5$. According to the distribution of the observation data, the initial parameters of the nonlinear membership function are shown in Table 3.

Table 1. Referential values of fault indicators.

	L	M	H
D_0	-8.32×10^{-4}	-8.01×10^{-4}	-7.69×10^{-4}
D_1	6.8×10^{-2}	7.1×10^{-2}	7.5×10^{-2}
D_2	2.36×10^{-2}	2.4×10^{-2}	2.43×10^{-2}

Table 2. The rules of initial BRB.

No.	θ_l	$D_0 \wedge D_1 \wedge D_2$	Consequent $\{B, M, S\}$	No.	θ_l	$D_0 \wedge D_1 \wedge D_2$	Consequent $\{B, M, S\}$
1	1	$L \wedge L \wedge L$	{0 0 1}	15	1	$M \wedge M \wedge H$	{0.4 0.5 0.1}
2	1	$L \wedge L \wedge M$	{0 0.1 0.9}	16	1	$M \wedge H \wedge L$	{0.4 0.6 0}
3	1	$L \wedge L \wedge H$	{0.1 0.1 0.8}	17	1	$M \wedge H \wedge M$	{0.5 0.3 0.2}
4	1	$L \wedge M \wedge L$	{0.1 0.2 0.7}	18	1	$M \wedge H \wedge H$	{0.5 0.4 0.1}
5	1	$L \wedge M \wedge M$	{0.1 0.3 0.6}	19	1	$H \wedge L \wedge L$	{0.4 0.3 0.3}
6	1	$L \wedge M \wedge H$	{0 0.4 0.6}	20	1	$H \wedge L \wedge M$	{0.5 0.3 0.2}
7	1	$L \wedge H \wedge L$	{0.1 0.4 0.5}	21	1	$H \wedge L \wedge H$	{0.6 0.2 0.2}
8	1	$L \wedge H \wedge M$	{0.2 0.2 0.6}	22	1	$H \wedge M \wedge L$	{0.6 0.3 0.1}
9	1	$L \wedge H \wedge H$	{0.2 0.3 0.5}	23	1	$H \wedge M \wedge M$	{0.6 0.2 0.2}
10	1	$M \wedge L \wedge L$	{0 0.5 0.5}	24	1	$H \wedge M \wedge H$	{0.7 0.2 0.1}
11	1	$M \wedge L \wedge M$	{0.1 0.6 0.3}	25	1	$H \wedge H \wedge L$	{0.8 0.2 0}
12	1	$M \wedge L \wedge H$	{0.2 0.5 0.3}	26	1	$H \wedge H \wedge M$	{0.9 0.1 0}
13	1	$M \wedge M \wedge L$	{0.2 0.6 0.2}	27	1	$H \wedge H \wedge H$	{1 0 0}
14	1	$M \wedge M \wedge M$	{0.3 0.6 0.1}				

Table 3. Initial parameters of the nonlinear membership function.

s_{ij}	$j=1$	$j=2$
$i = 1$	0.5	0.7
$i = 2$	2	1.5
$i = 3$	1	2

Since it is difficult for the initially constructed BRB to achieve ideal modeling accuracy, observation data are required to optimize the model. For 180 groups of observation data, 30% are randomly selected as the training set and the rest as the test set to reflect the modeling ability of BRB in small sample problems. The gradient descent method in Section 4.3 is used as the optimization engine. The optimized model parameters are shown in Table 4. In addition, optimized parameters of the nonlinear membership function are shown in Table 5. The fault diagnosis results of optimized BRB and initial BRB are shown in Figure 7.

Table 4. The rules of optimized BRB.

No.	θ_l	$D_0 \wedge D_1 \wedge D_2$	Consequent $\{B, M, S\}$	No.	θ_l	$D_0 \wedge D_1 \wedge D_2$	Consequent $\{B, M, S\}$
1	0.81	$L \wedge L \wedge L$	{0.13 0.06 0.81}	15	0.80	$M \wedge M \wedge H$	{0.44 0.46 0.10}
2	0.91	$L \wedge L \wedge M$	{0.13 0.16 0.71}	16	0.14	$M \wedge H \wedge L$	{0.35 0.61 0.04}
3	0.13	$L \wedge L \wedge H$	{0.10 0.11 0.79}	17	0.42	$M \wedge H \wedge M$	{0.41 0.33 0.26}
4	0.91	$L \wedge M \wedge L$	{0.20 0.25 0.55}	18	0.92	$M \wedge H \wedge H$	{0.54 0.34 0.12}
5	0.63	$L \wedge M \wedge M$	{0.18 0.25 0.57}	19	0.79	$H \wedge L \wedge L$	{0.39 0.28 0.33}
6	0.10	$L \wedge M \wedge H$	{0.02 0.41 0.58}	20	0.74	$H \wedge L \wedge M$	{0.52 0.29 0.19}
7	0.28	$L \wedge H \wedge L$	{0.13 0.34 0.54}	21	0.66	$H \wedge L \wedge H$	{0.51 0.23 0.27}
8	0.55	$L \wedge H \wedge M$	{0.26 0.19 0.55}	22	0.34	$H \wedge M \wedge L$	{0.50 0.36 0.14}
9	0.96	$L \wedge H \wedge H$	{0.27 0.32 0.41}	23	0.85	$H \wedge M \wedge M$	{0.57 0.26 0.18}
10	0.84	$M \wedge L \wedge L$	{0.13 0.43 0.43}	24	0.93	$H \wedge M \wedge H$	{0.69 0.20 0.12}
11	0.16	$M \wedge L \wedge M$	{0.10 0.52 0.38}	25	0.68	$H \wedge H \wedge L$	{0.67 0.24 0.09}
12	0.93	$M \wedge L \wedge H$	{0.25 0.44 0.31}	26	0.76	$H \wedge H \wedge M$	{0.76 0.17 0.07}
13	0.45	$M \wedge M \wedge L$	{0.30 0.46 0.24}	27	0.74	$H \wedge H \wedge H$	{0.92 0.03 0.06}
14	0.49	$M \wedge M \wedge M$	{0.33 0.57 0.11}				

Table 5. Optimized parameters of the nonlinear membership function.

s_{ij}	$j=1$	$j=2$
$i = 1$	0.65	0.84
$i = 2$	2.12	1.76
$i = 3$	1.05	2.41

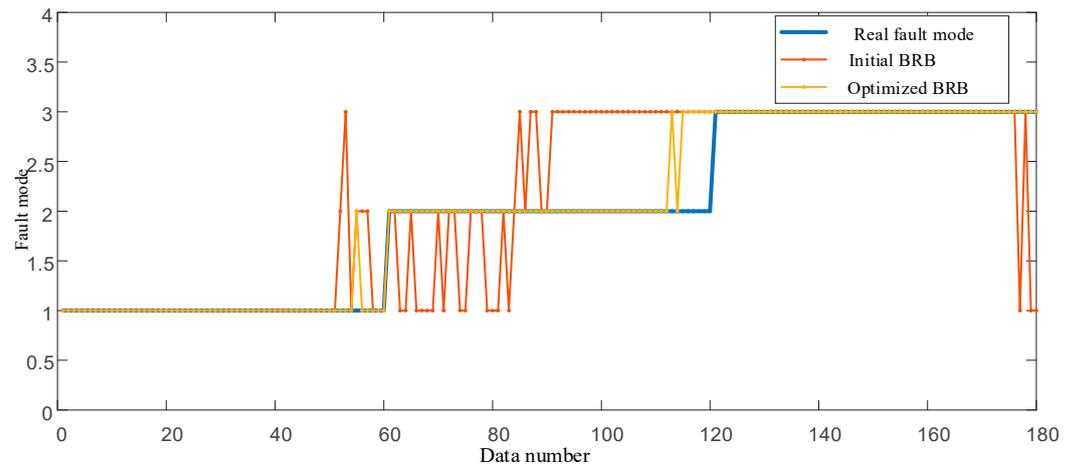


Figure 7. Comparison between initial BRB and optimized BRB.

4.3. Comparative Study

In order to fully verify the effectiveness of the model in this paper, comparative experiments are carried out in this section from the following two aspects, namely, the previous BRB model and data-driven models.

a. The previous BRB model

(1) The BRB model with triangular membership function, named BRB-t: the input information transformation function of this BRB adopts the triangular membership function in Equation (4). BRB-tri also uses the gradient descent method to optimize model parameters.

(2) The BRB model with Gaussian membership function, is named BRB-g: the optimization method of this model is the same as BRB-t.

Correspondingly, the BRB proposed in this paper is named BRB-n. The fault diagnosis results of three BRBs on 180 sets of observation data are shown in Figure 8. The accuracy of test data is shown in Table 6.

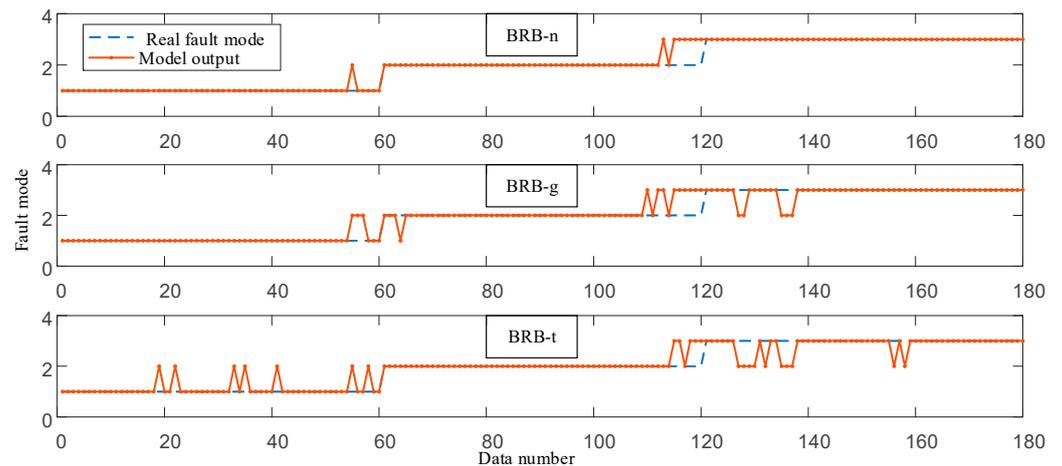


Figure 8. Comparison of diagnosis results of different BRBs.

Table 6. Comparison of modeling accuracy of different BRBs.

	Initial BRB	BRB-n	BRB-g	BRB-t
Diagnostic accuracy	65.68%	95.56%	90.74%	87.21%

b. Data-driven models

Data-driven fault diagnosis methods have been widely used. In this paper, Random forest (RF), Naive Bayes (NB) and K-nearest neighbor (KNN) models are used for comparative study.

(1) RF: RF is a type of powerful tree ensemble model [33]. Its basic model is the decision tree (DT). Compared with general DT, RF has a stronger generalization ability, so it is widely used in classification and regression problems.

(2) NB: NB is a nonparametric model based on the Bayesian theorem [34]. This model has no explicit learning process. Generally, it calculates the prior probability and the conditional probability directly from the training set and infers a posteriori probability.

(3) KNN: KNN is a lazy machine learning model [35], which means that it has no model training process. For the data to be predicted, the training data closest to this data are first obtained according to the defined distance formula. Then, their weighted averages or votes are calculated.

The hyper-parameters of these models are the default settings in the Python “sklearn” library. They are then adjusted by the “GridSearchCV” function. Their diagnostic results and accuracy are shown in Figure 9 and Table 7, respectively.

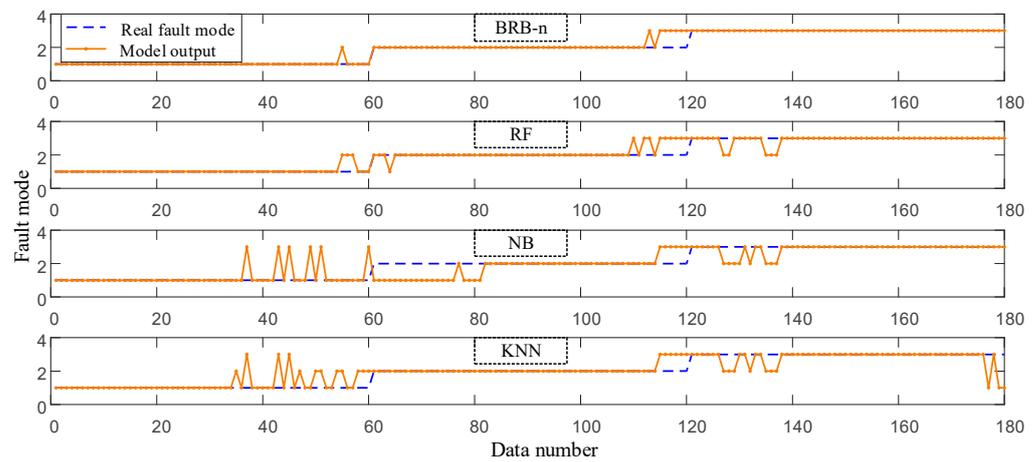


Figure 9. Comparison of diagnosis results of d data-driven models.

Table 7. Comparison of modeling accuracy of data-driven models.

	BRB-n	RF	NB	KNN
Diagnostic accuracy	95.56%	84.72%	76.30%	72.91%

c. Swarm intelligence algorithms

To illustrate the advantages of the gradient-based optimization algorithm proposed in this paper, three swarm intelligence algorithms, that is, DE, PSO and P-CMAES, are used to optimize the initial BRB model. Their parameter settings are the same as those in [23–25]. The optimized BRBs are named BRB-DE, BRB-PSO and BRB-PCMAES, respectively. For comparison, the BRB optimized by the proposed method is named BRB-GB. The accuracy of fault diagnosis of these models is shown in Table 8.

Table 8. Comparison of modeling accuracy of different optimization algorithms.

	BRB-DE	BRB-PSO	BRB-PCMAES	BRB-GB
Diagnostic accuracy	88.59%	89.41%	92.74%	95.56%

4.4. Analysis and Discussions

Firstly, the advantages of the BRB-based fault diagnosis model are described by comparing it with data-driven models. In this paper, 30% of samples of the dataset are selected as the training set to train a wide variety of fault diagnosis models. Under this circumstance, the drawbacks of the data-driven model are revealed. Since the training set may not comprehensively reflect the overall mapping relationship, these models almost fall into the problem of overfitting. Among them, with the advantage of a double random sampling of features and samples, the random forest model alleviates the problem to some extent and has the highest diagnostic accuracy among several data-driven models. By comparing Tables 5 and 6, it can be seen that after model optimization, the fault diagnosis accuracy of several BRBs are higher than that of data-driven models. Among them, the diagnostic accuracy of BRB-n has been improved by 12.8%, 25.24% and 31.07%, respectively. This reveals the advantage of expert knowledge in small sample modeling; that is, experts construct a generally correct but rough model by virtue of domain knowledge and experience. Then, the diagnosis accuracy of the model is further improved through the existing small sample dataset. On the other hand, as a “white-box” model, the BRB model provides explicit knowledge representation and reasoning compared with data-driven methods, enabling the results of fault diagnosis to be traceable and transparent. In order to further illustrate the interpretability of BRB, the fault diagnosis process of the first observation data $[-7.94 \times 10^{-4} \ 6.82 \times 10^{-2} \ 2.38 \times 10^{-2}]$ is shown in Figure 10.

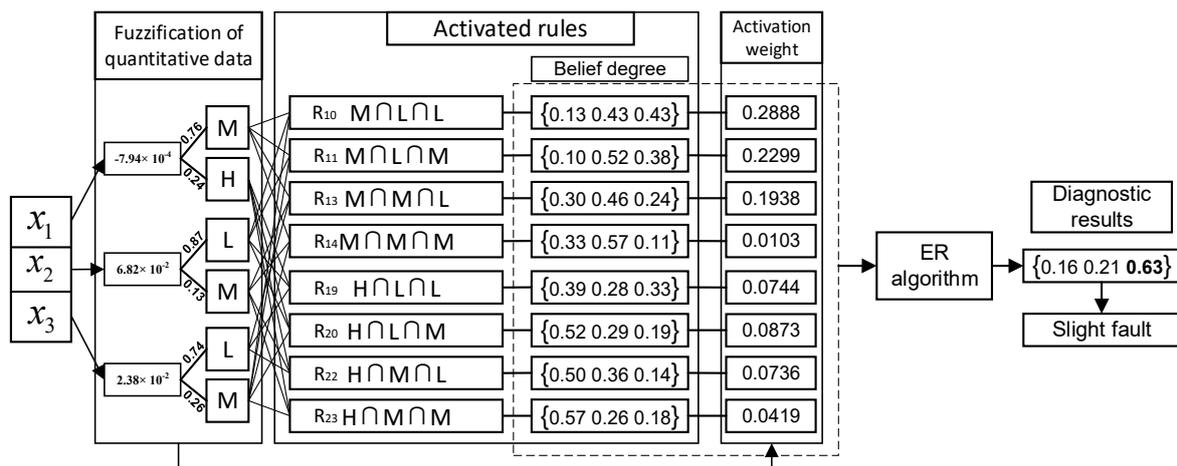


Figure 10. Visualization of fault diagnosis based on BRB.

Secondly, the advantages of the gradient-based optimization method are analyzed. On the one hand, the gradient-based optimization algorithm shows powerful model training ability. Compared with the other three swarm intelligence algorithms, the modeling accuracy of BRB-GB has improved by 7.87%, 6.44% and 2.95%, respectively. On the other hand, the proposed optimization algorithm can keep expert knowledge from being destroyed, which is reflected in that the optimized model parameters will not be significantly changed, but rather, fine-tuned. For simplicity, the belief degrees of the first rule in four BRBs are shown in Table 9. Table 9 shows that, due to random initialization, the distribution of belief degree of the BRB optimized by the swarm intelligence algorithm has seriously deviated from the initial judgment of experts, even if they can achieve good fault diagnosis accuracy. This will make the rules difficult to understand; when the referential value of the evaluation indicator is low, the fault degree is generally “slight” according to common sense and experience. When these BRBs are used for fault diagnosis, the interpretability of the diagnosis results will be weakened.

Table 9. The distribution of belief degree of the first rule in the four BRBs.

Model	Belief Degree of the First Rule	The Most Supported Grade
Initial BRB	{0 0 1}	S
BRB-GB	{0.13 0.06 0.81 }	S
BRB-DE	{0.21 0.52 0.27}	M
BRB-PSO	{0.15 0.65 0.20}	M
BRB-PCMAES	{ 0.43 0.28 0.29}	B

Finally, the importance of the nonlinear membership function is analyzed. First, the uneven distribution of observation data will affect the accuracy of the fault diagnosis model. It can be seen from Table 6 that the diagnosis accuracy of BRB is improved after considering the adaptive membership function. This is easy to understand since the original triangular membership function cannot reflect the diversity of different data distributions, leading to errors in the fuzzification of input information. Therefore, it is necessary to consider the adaptive membership function when building a fault diagnosis model based on BRB. Then, compared with the other two BRBs, the diagnostic accuracy of BRB-n is increased by 5.31% and 9.57%, respectively. It is worth noting that the Gaussian membership function has more adjustment parameters than the nonlinear membership function. For BRB-g and BRB-n in this paper, the optimization parameters are 123 and 117, respectively. With the increase of the number of the referential grade, the difference in the number of parameters will continue to expand. However, the nonlinear membership function can show better performance. This is because, compared with the Gaussian membership function, the nonlinear membership function can adapt to a wider range of data distributions and can more precisely conduct the fuzzification of quantitative data.

In order to further verify the role of information transformation of nonlinear membership function, part of the observation data of D_2 is shown in Figure 11. The observation data are concentrated on H_{32} in the interval $[a_{32}, a_{33}]$. Therefore, in the entire dataset, the red-marked areas are more subordinate in H_{33} . In other words, this point should belong to the referential grade “high” in the entire dataset to a greater extent. According to the optimized BRB reasoning process, the membership degrees of H_{32} and H_{33} are 0.18 and 0.82, respectively, which is consistent with the real distribution of the dataset. Therefore, when the nonlinear membership function is used to transform the input information, BRB can achieve a more ideal modeling accuracy.

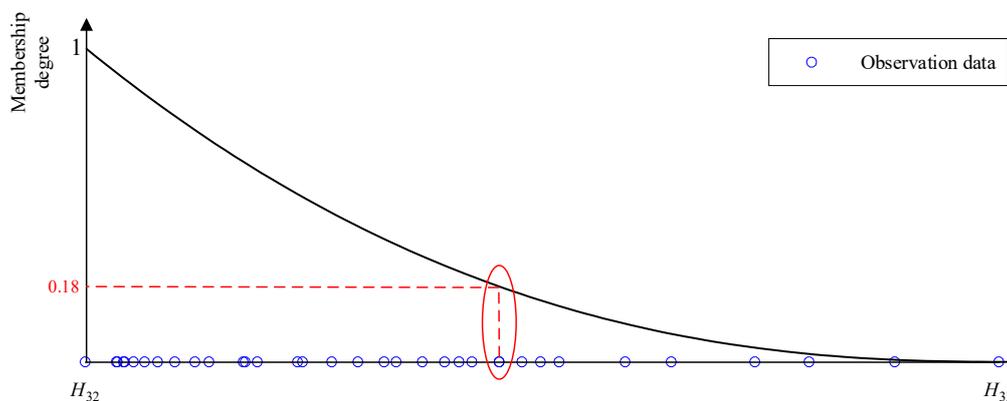


Figure 11. Membership degree of data with non-uniform distribution.

5. Conclusions

In this paper, a new fault diagnosis model based on BRB has been proposed. In order to address the distribution of different data, an adaptive nonlinear membership function has been proposed to conduct the fuzzification of quantitative data. Since the parameters of the membership function initially determined by experts may not be accurate in the

new BRB model, a new parameter optimization model considering the parameters of the membership function has been proposed with the aid of by the gradient descent algorithm. Finally, the proposed model is verified by a laser gyro fault diagnosis case.

In summary, the proposed method has two advantages: firstly, in the transformation of input information, the limitations of the triangular membership function in the fuzzification of non-uniformly distributed observation data are considered for the first time and an adaptive nonlinear membership function is designed. This function can adapt to the distribution of various data and improve the accuracy of information transformation. Secondly, considering the subjectivity and ignorance of experts in determining the parameters of the model, the parameters of the membership function are added to the optimization model; the gradient descent method is used to optimize the fault diagnosis model, enabling expert knowledge to not be destroyed and improving modeling accuracy.

Furthermore, the model optimization of BRB is a non-convex optimization problem, which means that the traditional gradient descent method may fall into the local optimal value. Therefore, it is interesting to get higher modeling accuracy by jumping out of the local optimal value. This issue will be considered in the future.

Author Contributions: Conceptualization, Z.L. and Z.Z.; methodology, X.Z.; validation, X.H.; formal analysis, Z.L.; investigation, Z.F.; resources, C.H.; data curation, Z.L.; writing—original draft preparation, Z.L.; writing—review and editing, Z.L.; visualization, Z.F.; supervision, Z.Z.; project administration, Z.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported in part by the Natural Science Foundation of China under Grants 61833016 and 6227814, the Shaanxi Outstanding Youth Science Foundation under Grant 2020JC-34, and the Shaanxi Science and Technology Innovation Team under Grant No. 2022TD-24.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Hu, Y.; Miao, X.; Si, Y.; Pan, E.; Zio, E. Prognostics and health management: A review from the perspectives of design, development and decision. *Reliab. Eng. Syst. Saf.* **2022**, *217*, 108063. [[CrossRef](#)]
2. Lee, J.; Wu, F.; Zhao, W.; Ghaffari, M.; Liao, L.; Siegel, D. Prognostics and health management design for rotary machinery systems—Reviews, methodology and applications. *Mech. Syst. Signal Process.* **2014**, *42*, 314–334. [[CrossRef](#)]
3. Chen, H.T.; Jiang, B. A review of fault detection and diagnosis for the traction system in high-speed trains. *IEEE Trans. Intell. Transp. Syst.* **2019**, *21*, 450–465. [[CrossRef](#)]
4. Cheng, C.; Wang, J.; Chen, H.; Chen, Z.; Luo, H.; Xie, P. A review of intelligent fault diagnosis for high-speed trains: Qualitative approaches. *Entropy* **2020**, *23*, 1. [[CrossRef](#)] [[PubMed](#)]
5. Zhang, Q.H. Adaptive Kalman filter for actuator fault diagnosis. *Automatica* **2018**, *93*, 333–342. [[CrossRef](#)]
6. Zhou, S.; Chen, Z.; Lin, T. Lithium-Ion Battery Cell Open Circuit Fault Diagnostics: Methods, Analysis, and Comparison. *IEEE Trans. Power Electron.* **2022**, *38*, 2493–2505. [[CrossRef](#)]
7. Ruijters, E.; Stoelinga, M. Fault tree analysis: A survey of the state-of-the-art in modeling, analysis and tools. *Comput. Sci. Rev.* **2015**, *15*, 29–62. [[CrossRef](#)]
8. Liu, Y.; Ni, W.; Ge, Z. Fuzzy decision fusion system for fault classification with analytic hierarchy process approach. *Chemom. Intell. Lab. Syst.* **2017**, *166*, 61–68. [[CrossRef](#)]
9. Chen, H.; Jiang, B.; Ding, S.X.; Huang, B. Data-driven fault diagnosis for traction systems in high-speed trains: A survey, challenges, and perspectives. *IEEE Trans. Intell. Transp. Syst.* **2022**, *23*, 1700–1716. [[CrossRef](#)]
10. Tran, V.T.; Yang, B.-S.; Oh, M.-S.; Tan, A.C.C. Fault diagnosis of induction motor based on decision trees and adaptive neuro-fuzzy inference. *Expert Syst. Appl.* **2009**, *36*, 1840–1849. [[CrossRef](#)]
11. Jia, P.; Wang, C.; Zhou, F.; Hu, X. Trend Feature Consistency Guided Deep Learning Method for Minor Fault Diagnosis. *Entropy* **2023**, *25*, 242.
12. Yang, J.B.; Liu, J.; Wang, J.; Sii, H.; Wang, H.W. Belief rule based inference methodology using the evidential reasoning approach—RIMER. *IEEE Trans. Syst. Man Cybern.-Part A* **2006**, *36*, 266–285. [[CrossRef](#)]
13. Cao, Y.; Zhou, Z.; Hu, C.; He, W.; Tang, S. On the interpretability of belief rule-based expert systems. *IEEE Trans. Fuzzy Syst.* **2020**, *29*, 3489–3503. [[CrossRef](#)]
14. Jiao, L.M.; Denoeux, T.; Pan, Q. A hybrid belief rule-based classification system based on uncertain training data and expert knowledge. *IEEE Trans. Syst. Man Cybern. Syst* **2015**, *46*, 1711–1723. [[CrossRef](#)]

15. Zhou, Z.J.; Hu, G.Y.; Hu, C.H.; Wen, C.L.; Chang, L.L. A survey of belief rule-based expert system. *IEEE Trans. Syst. Man Cybern. Syst.* **2019**, *51*, 4944–4958. [[CrossRef](#)]
16. Xu, X.J.; Yan, X.; Sheng, C.; Yuan, C.; Xu, D.; Yang, J.B. A belief rule-based expert system for fault diagnosis of marine diesel engines. *IEEE Trans. Syst. Man Cybern. Syst.* **2017**, *50*, 656–672. [[CrossRef](#)]
17. Feng, Z.C.; Zhou, Z.J.; Hu, C.H.; Yin, X.J.; Hu, G.Y.; Zhao, F.J. Fault diagnosis based on belief rule base with considering attribute correlation. *IEEE Access* **2017**, *6*, 2055–2067. [[CrossRef](#)]
18. Zhang, C.C.; Zhou, Z.J.; Tang, S.W.; Chen, L.Y.; Zhang, P. BR-FRL: A Belief Rule-Based Fault Recognition and Location Model for Bus Network Systems. *IEEE Trans. Instrum. Meas.* **2022**, *71*, 1–12. [[CrossRef](#)]
19. Li, C.; Shen, Q.; Wang, L.X.; Qin, W.; Xie, M.M. A New Adaptive Interpretable Fault Diagnosis Model for Complex System Based on Belief Rule Base. *IEEE Trans. Instrum. Meas.* **2022**, *71*, 1–11. [[CrossRef](#)]
20. Chen, M.L.; Zhou, Z.J.; Zhang, B.C.; Hu, G.Y.; Cao, Y. A novel combination belief rule base model for mechanical equipment fault diagnosis. *Chin. J. Aeronaut.* **2021**, *35*, 158–178. [[CrossRef](#)]
21. Ming, Z.C.; Zhou, Z.J.; Cao, Y.; Tang, S.W.; Chen, Y.; Han, X.X.; He, W. A new interpretable fault diagnosis method based on belief rule base and probability table. *Chin. J. Aeronaut.* **2022**. [[CrossRef](#)]
22. Wang, Z.Y.; Li, S.H.; He, W.; Yang, R.H.; Feng, Z.C.; Sun, G.W. A New Topology-Switching Strategy for Fault Diagnosis of Multi-Agent Systems Based on Belief Rule Base. *Entropy* **2022**, *24*, 1591. [[CrossRef](#)] [[PubMed](#)]
23. Yang, J.B. Rule and utility based evidential reasoning approach for multiattribute decision analysis under uncertainties. *Eur. J. Oper. Res.* **2001**, *131*, 31–61. [[CrossRef](#)]
24. Mas, M.; Monserrat, M.; Torrens, J.; Trillas, E. A survey on fuzzy implication functions. *IEEE Trans. Fuzzy Syst.* **2007**, *15*, 1107–1121. [[CrossRef](#)]
25. Liu, J.; Yang, J.; Ruan, D.; Martinez, L.; Wang, J. Self-tuning of fuzzy belief rule bases for engineering system safety analysis. *Ann. Oper. Res.* **2008**, *163*, 143–168. [[CrossRef](#)]
26. Li, G.L.; Zhou, Z.J.; Hu, C.H.; Chang, L.L.; Zhou, Z.G.; Zhao, F.J. A new safety assessment model for complex system based on the conditional generalized minimum variance and the belief rule base. *Saf. Sci.* **2017**, *93*, 108–120. [[CrossRef](#)]
27. Qian, B.; Wang, Q.Q.; Hu, R.; Zhou, Z.J.; Yu, C.Q.; Zhou, Z.G. An effective soft computing technology based on belief-rule-base and particle swarm optimization for tipping paper permeability measurement. *J. Ambient. Intell. Humaniz. Comput.* **2019**, *10*, 841–850. [[CrossRef](#)]
28. Zhou, Z.J.; Hu, G.Y.; Zhang, B.C.; Hu, C.H.; Zhou, Z.G.; Qiao, P.L. A model for hidden behavior prediction of complex systems based on belief rule base and power set. *IEEE Trans. Syst. Man Cybern. Syst.* **2017**, *48*, 1649–1655. [[CrossRef](#)]
29. Chen, Y.W.; Yang, J.B.; Xu, D.L.; Yang, S.L. On the inference and approximation properties of belief rule based systems. *Inf. Sci.* **2013**, *234*, 121–135. [[CrossRef](#)]
30. Wang, Y.M.; Yang, J.; Xu, D.L. Environmental impact assessment using the evidential reasoning approach. *Eur. J. Oper. Res.* **2006**, *174*, 1885–1913. [[CrossRef](#)]
31. Yang, J.B.; Liu, J.; Xu, D.L.; Wang, H.W. Optimization models for training belief-rule-based systems. *IEEE Trans. Syst. Man Cybern. -Part A Syst. Hum.* **2007**, *37*, 569–585. [[CrossRef](#)]
32. Feng, Z.C.; Zhou, Z.J.; Hu, C.H.; Ban, X.J.; Hu, G.Y. A safety assessment model based on belief rule base with new optimization method. *Reliab. Eng. Syst. Saf.* **2020**, *203*, 107055. [[CrossRef](#)]
33. Yang, B.S.; Di, X.; Han, T. Random forests classifier for machine fault diagnosis. *J. Mech. Sci. Technol.* **2008**, *22*, 1716–1725. [[CrossRef](#)]
34. Cai, B.; Huang, L.; Xie, M. Bayesian networks in fault diagnosis. *IEEE Trans. Ind. Inform.* **2017**, *13*, 2227–2240. [[CrossRef](#)]
35. He, Q.P.; Wang, J. Fault detection using the k-nearest neighbor rule for semiconductor manufacturing processes. *IEEE Trans. Semicond. Manuf.* **2007**, *20*, 345–354. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.