Belief Rule Base Structure and Parameter Joint Optimization Under Disjunctive Assumption for Nonlinear Complex System Modeling

Lei-Lei Chang, Zhi-Jie Zhou, Yu-Wang Chen, Tian-Jun Liao, Yu Hu, and Long-Hao Yang

***Abstract*—Nonlinear complex system modeling has drawn attention from diverse fields and many approaches have been developed. Among those approaches, the advantages of the belief rule base (BRB) expert system have been shown for managing multiple types of information under uncertainty and modeling the nonlinearity present in many theoretical and practical complex systems. However, two challenges still need to be addressed. First, BRB needs to be downsized to conserve modeling and computational effort. For this challenge, a new disjunctive assumption is applied, which can significantly downsize BRB while maintaining its completeness. Second, the structure and parameters of BRB need to be jointly optimized. For this challenge, a new Akaike information criterion (AIC)-based optimization objective is derived to represent both modeling accuracy and modeling complexity. Moreover, a joint bi-level optimization model with an AIC-based objective is constructed for the BRB structure and parameters, and a bi-level optimization algorithm is proposed. Three evolutionary algorithms, namely, the genetic algorithm, particle swarm optimization algorithm, and differential evolutionary algorithm, are tested in a comparative fashion to determine the best fit for the optimization engine. The results of two practical case studies show that the joint optimization approach can identify an optimal configuration for both its structure and parameters, which is referred to as the best decision structure in this paper.**

Manuscript received September 18, 2016; revised January 13, 2017; accepted February 19, 2017. The work of L.-L. Chang was supported in part by the National Natural Science Foundation of China (NSFC) under Grant 71501182, Grant 71571185, Grant 71601180, Grant 71671186, and Grant 61403404, and in part by the open funding programme of Joint Laboratory of Flight Vehicle Ocean-based Measurement and Control under Grant FOM2015OF017. The work of Z.-J. Zhou was supported in part by the NSFC under Grant 61370031 and Grant 61374138, in part by the Post-Doctoral Science Foundation of China under Grant 2015M570847, and in part by the Natural Science Foundation of Shaanxi Province under Grant 2015JM6354. The work of T.-J. Liao was supported by the NSFC under Grant 71401167. The work of Y. Hu was supported by the China Aviation Science Foundation under Grant 201605U8002. This paper was recommended by Associate Editor J.-H. Chou. *(Corresponding authors: Lei-Lei Chang; Yu Hu.)*

L.-L. Chang, Z.-J. Zhou, and Y. Hu are with the High-Tech Institute of Xi'an, Xi'an 710025, China (e-mail: leileichang@hotmail.com; zhouzj04@mails.tsinghua.edu.cn; huyu1222@163.net).

Y.-W. Chen is with the Decision and Cognitive Science Research Centre, Manchester Business School, University of Manchester, Manchester M15 6PB, U.K. (e-mail: yu-wang.chen@manchester.ac.uk).

T.-J. Liao is with the State Key Laboratory of Complex System Simulation, Beijing Institute of System Engineering, Beijing 100101, China (e-mail: liaotianjun@gmail.net).

L.-H. Yang is with the Department of Management, Fuzhou University, Fuzhou 350116, China (e-mail: more026@hotmail.edu).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TSMC.2017.2678607

***Index Terms*—Belief rule base (BRB), complex system modeling, disjunctive assumption, joint optimization.**

I. Introduction

NONLINEARITY is natural in many complex systems, and efficiently modeling and analyzing nonlinearity is of great significance for better understanding and predicting complex system behaviors [14], [25]. Multiple researchers in diverse fields have constructed various models/approaches. Based on the type of primary functioning factors, they can be categorized as qualitative approaches, e.g., the Delphi approach [28] or AHP [29], and quantitative approaches, e.g., experimental and statistical approaches [36]. Qualitative approaches are mainly used under conditions when there is no clear understanding of the analytical correlations between the complex systems and/or experts' experience and knowledge are involved. In contrast, the results of quantitative approaches are considered to be more accurate, because they are based on experiments, statistics, and mathematical deductions. The main disadvantage of subjective approaches is that they can be biased due to human preferences and/or conflicts of interests. On the other hand, quantitative approaches are very expensive, and sometimes experiments cannot be repeated a large number of times. Many additional approaches with both qualitative and quantitative characteristics have been proposed, including expert-system-based approaches [12], [22], failure-oriented approaches [13], [15], simulation-based approaches [32], [44], characteristic-based approaches [33], hybrid approaches [37], etc.

Among those approaches, rule-based systems have been validated to be efficient in solving many complex system modeling problems, especially fuzzy-rule-based systems [43] and belief-rule-based systems [16], [41]. However, it has also been argued that the selection of certain parameters, particularly the number of fuzzy partitions and the number of rules, can influence modeling performance [7], [17]. Comparatively, belief rule base (BRB) has the advantage of generalizing the conventional probability distribution by allowing inexact reasoning [16] and dealing with uncertain information which is different from ignorance and equal likelihood [42]. In recent years, BRB has been successfully applied to solve problems and meet challenges in multiple fields [4], [20], [23], [45].

However, two challenges need to be addressed. The first is how to downsize BRB and the second is how to jointly optimize the structure and parameters of BRB.

A. First Challenge: How to Downsize BRB

Although neither the conjunctive nor disjunctive assumption was compulsorily requested in the original BRB definition [39], the conjunctive assumption was applied in most studies; this assumption requires covering every possible combination of all referenced values of all attributes [8], [41]. There are corresponding rules to be activated for any input. In other words, the traditional BRB is complete. However, a traditional BRB under the conjunctive assumption faces the combinatorial explosion problem. For example, if a BRB has three/four/five attributes, and each attribute has three referenced values, then there would be $3^3 = 27$, $3^4 = 81$, and $3^5 = 243$ rules, respectively. In a word, the size of BRB (the number of rules in the BRB) would grow exponentially along with more attributes and/or more reference values of the attributes. It would be considerably difficult to gather enough information to quantify and optimize BRB parameters with multiple rules.

Several researchers have made many important attempts to solve this problem. To reduce the number of attributes, Chang *et al.* [7] proposed a structure learning approach for BRB, in which the initial five attributes were reduced to three and the number of rules in the BRB was reduced from 243 to 27 (an 88.89% reduction). Zhou *et al.* [48] used "statistical utility" as the criterion to filter rules, by excluding rules with a relatively low probability. However, this method results in an incomplete BRB. Wang *et al.* [34] proposed an extended BRB approach to determine the best fit number of rules in BRB. Another method optimizes the number of the referenced values of the attributes, which is referred to as the parameter learning for BRB. Chang *et al.* [5] used the differential evolutionary (DE) algorithm as the optimization engine for BRB parameter learning. In other related studies on the parameter learning for BRB, including the online BRB parameter learning approach [47], [49], the number of referenced values of the attributes has not been considered as a parameter to be optimized.

To summarize, present BRB-related studies are still based on the traditional conjunctive assumption, which requires transversely constructing the BRB. From this point of view, the size of BRB cannot be significantly reduced. In this paper, a new disjunctive assumption replaces the conventional conjunctive assumption [8]. This paper also discusses certain characteristics of the disjunctive assumption, and demonstrates that the BRB constructed under the disjunctive assumption would still be complete.

As previously addressed, BRB structure and parameter learning approaches have been studied separately, which leads to the second challenge: how to jointly optimize BRB.

B. Second Challenge: How to Jointly Optimize BRB

In many approaches and models, e.g., fuzzy set theory, BRB, and neural networks, modeling accuracy has been the most important objective [11], [38], [46]–[49]. However, modeling complexity should also be taken into consideration, because it directly affects not only understandability but also the learning process. Unfortunately, modeling accuracy and modeling complexity are not consistent with each other under most circumstances. On one hand, higher modeling accuracy requires more rules, so that the local nonlinearity of a complex system can be accurately modeled. On the other hand, more rules results in higher complexity, because there is a more complex structure, more rules, and more parameters to understand and optimize. Therefore, modeling accuracy and modeling complexity need to be balanced via a jointly optimization of the structure and parameters of BRB.

For the second challenge, the first generic BRB parameter learning framework and corresponding optimization model were proposed by [40]. Xu *et al.* [38] further proposed a BRB-inferring and training approach, and applied it to pipeline leak detection. Chen *et al.* [11] proposed an adaptive training approach to optimize BRB parameters. Zhou *et al.* [47], [49] argued that these approaches were offline, and proposed the first online updating approach. The online approach also integrated expert knowledge in order to gather more information to better constructing BRB. Additionally, this online approach did not require a complete set of data to be collected before training the BRB. Later, Chen *et al.* [10] proposed a more comprehensive learning and inference framework for BRB. The above studies were all performed on BRB parameter learning. Chang *et al.* [7] conducted the first BRB structure learning approach.

To conclude, modeling accuracy has so far been the sole objective for BRB structure and parameter learning, which have thus far been conducted separately. However, the derived model with the highest modeling accuracy may also have high modeling complexity, and would require excessive optimization and computation efforts. This paper proposes a joint optimization approach for BRB structure and parameter learning.

After the joint optimization of BRB structure and parameters, the optimal BRB structure and parameters are derived, which is referred to as the best decision structure of BRB in this paper, denoting that the derived BRB has an optimal configuration in terms of both its structure and parameters.

Note that Savan *et al.* [31] conducted a study that also employed Akaike information criterion (AIC) as its objective for BRB structure validation. The main differences between the two studies, which are also the main contributions of this paper are that: 1) the disjunctive assumption is considered for downsizing BRB, while [31] used the conjunctive assumption and 2) a bi-level optimization model is constructed, and a corresponding bi-level optimization algorithm is developed for reducing computational burden.

The remainder of this paper is organized as follows. The basics of BRB are introduced in Section II. The disjunctive assumption is introduced in Section III. The BRB joint optimization model and optimization algorithm are presented in Sections IV and V, respectively. Two practical cases are studied in Section VI to validate the efficiency of the BRB joint optimization approach. Section VII presents the conclusion.

II. Basics and Challenges of BRB

A. BRB Basics

BRB is composed of multiple belief rules in the same belief structure, in which the kth rule is described as [41]

$$\begin{aligned} R_k:\ &\text{if } \left(x_1 \text{ is } A_1^k\right) \wedge \left(x_2 \text{ is } A_2^k\right) \wedge \cdots \wedge \left(x_M \text{ is } A_M^k\right) \\ &\text{then } \left\{\left(D_1, \beta_{1,k}\right), \ldots, \left(D_N, \beta_{N,k}\right)\right\} \\ &\text{with rule weight } \theta_k \end{aligned} \tag{1}$$

where $x_m (m = 1, \ldots, M)$ denotes the mth attribute, $A_m^k (m = 1, \ldots, M; k = 1, \ldots, K)$ denotes the referenced values of the mth attribute in the kth rule, M denotes the number of attributes, $\beta_{n,k} (n = 1, \ldots, N)$ denotes the belief for the nth scale, D_n, in the conclusion part of the kth rule. N denotes the number of scales.

In the antecedent attribute part, different types of information under uncertainty could be taken as the input, and then transformed in the same belief structure. "$\wedge$" in (1) denotes that the BRB is constructed under the conjunctive assumption.

In the conclusion part, there are multiple scales, and each scale has a corresponding belief denoting how confident this rule is classified as this scale. Yang and Xu [42] demonstrated that the belief in the conclusion part is, in essence, the probability, which clarifies the ambiguity with respect to the probability theory.

The rule weight denotes the confidence in the corresponding rule. The rule weight could be derived either by experts or by statistical methods. The rule weight should be taken into consideration, with the matching degrees between the input and the activated rules forming the activated rule weight for further inference.

With the belief structure, BRB can handle different types of information under uncertainty by integrating both expert knowledge and historic data. Moreover, the inference process is open to decision makers and stakeholders, and the inference results are easy to interpret, and can be used to trace the source of influencing factors.

B. Advantages and Challenges of the Conjunctive Assumption

The biggest advantage of the conjunctive assumption for BRB construction is that it ensures BRB's completeness.

1) *BRB Is Complete:* The constructed BRB can handle any type of input.

In order to explain this advantage, we first introduce the concept of the BRB input set based on [10].

Definition 1: Let S be the BRB input set, M be the cardinality of BRB attributes, x_m be the mth attribute with a value range of A_m. S can be expressed as $S = \{(x_1, \ldots, x_m, \ldots, x_M) | x_m \in A_m, m = 1, \ldots, M\}$.

Fig. 1 gives an example of the input set of a BRB with two attributes: A and B. Each attribute has three referenced values: A_1, A_2, A_3, and B_1, B_2, B_3, so the BRB input set is a plane with nine intersected points (A_1B_3, A_2B_3, A_3B_3, A_1B_2, A_2B_2, A_3B_2, A_1B_1, A_2B_1, and A_3B_1).

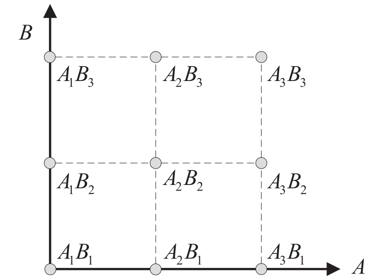


Fig. 1. BRB input set with two attributes and three referenced values for each attribute under the conjunctive assumption.

For the BRB input set, the concept of "BRB completeness" is given as follows.

1) *The BRB Is Complete:* All inputs are within the BRB input set, and any rule activated by the input is also within the BRB input set.

The completeness of a BRB input set is very important, because only so can BRB handle any input, regardless of the referenced values of the attributes.

Based on the concept of the BRB input set, the BRB construction and rule activation procedures are explained as follows.

1) The BRB is transversely constructed. Any intersected point in the BRB input set is the projection of the referenced attribute values (a point in the attribute). Therefore, the number of intersected points in the BRB input set (the number of rules in the BRB) is related to both the number of attributes and the number of referenced values of each attribute, as in

$$\text{num}_{\text{rules}} = \prod_{i=1}^{M} \text{num}_i \tag{2}$$

where $\text{num}_{\text{rules}}$ is the number of rules in the BRB, num_i is the number of referenced values for the ith attribute, and M is the number of attributes.

2) The points are activated by neighbor-intersected points in the BRB input set. Therefore, the number of activated rules $\text{num}_{\text{activated}}$ is only relevant with respect to the number of attributes in

$$\text{num}_{\text{activated}} = 2^M. \tag{3}$$

As we have established, a BRB input set under the conjunctive assumption has the advantage completeness. However, it also has the disadvantage of the combinatorial explosion problem: the number of (activated) rules in the BRB grows exponentially with more attributes, as explained in (2) and (3).

Challenge 1: Replacing the traditional conjunctive assumption requires a new assumption for constructing a BRB that is both downsized and complete. This challenge will be addressed in Section III.

C. Challenges of BRB Structure and Parameter Learning

So far, studies on BRB structure and parameter learning have focused on either the structure or parameters of BRB, instead of considering both. As explained in part A of Fig. 2,

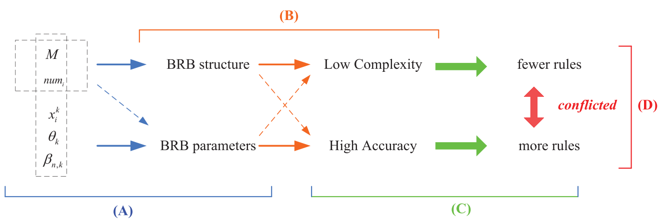


Fig. 2. Challenges of BRB structure and parameters learning.

the number of attributes and the number of referenced values of the attributes co-determine the structure and size of the BRB. The remaining three variables are used in BRB parameter learning (although it could be argued that all five variables are BRB parameters).

Refer to part B in Fig. 2. Naturally, modeling accuracy is the common goal for both BRB structure and parameter learning. However, modeling complexity should also be taken into consideration from a practical perspective, because BRB becomes infeasible for models with very high complexity, which makes human understanding and model optimization difficult. Thus, the optimized model produced by BRB structure and parameter learning should have both lower complexity and higher accuracy.

As shown in part C of Fig. 2, low complexity requires fewer parameters, and thus, fewer rules. On the other hand, high accuracy requires more rules, so that even the local nonlinearity of a complex system can be accurately modeled. Apparently, the two objectives conflict with one another, as shown in part D of Fig. 2.

Challenge 2: Representing both modeling accuracy and complexity requires the further construction of an optimization model. A corresponding optimization algorithm must be proposed for BRB joint optimization, which will be addressed in Sections IV and V.

III. Disjunctive Assumption for BRB Construction

A. Disjunctive Assumption

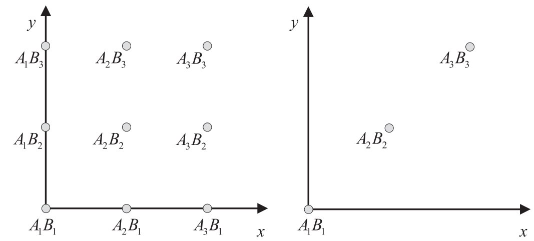
The disjunctive assumption for BRB is first introduced for the classification problem in [8]. The kth rule under the disjunctive assumption is given in

$$\begin{aligned} R_k:\ &\text{if } \left(x_1 \text{ is } A_1^k\right) \vee \left(x_2 \text{ is } A_2^k\right) \vee \cdots \vee \left(x_M \text{ is } A_M^k\right) \\ &\text{then } \left\{\left(D_1, \beta_{1,k}\right), \ldots, \left(D_N, \beta_{N,k}\right)\right\} \\ &\text{with rule weright } \theta_k \end{aligned} \tag{4}$$

where "$\vee$" in (1) denotes the disjunctive assumption, which means that a rule is activated if either attribute condition is met. Therefore, the weight of the activated rule is calculated by

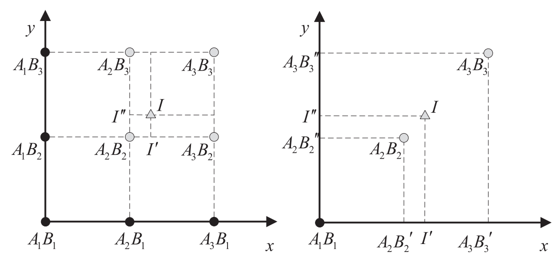
$$w_k = \frac{\theta_k \sum_{i=1}^{M} \alpha_i^k}{\sum_{l=1}^{L} \theta_l \sum_{i=1}^{M} \alpha_i^l} \tag{5}$$

where θ_k represents the initial weight of the kth rule, and α_i^k represents the matching degree between the input of the ith attribute in the kth rule.



(a) (b)

Fig. 3. BRB input set under the (a) conjunctive assumption and (b) disjunctive assumption.



(a) (b)

Fig. 4. BRB input set with two attributes and three reference values for each attribute under the (a) conjunctive and (b) disjunctive assumptions.

Fig. 3(a/b) shows the BRB input set under conjunctive and disjunctive assumptions. In Fig. 3(b), only three rules (points in the BRB input set) are required, because at least two rules are required to maintain the upper and lower bounds of each attribute.

The rule activation, matching degree calculation, and weight calculation procedures for the disjunctive assumption can be found in [8].

B. Characteristics of the Disjunctive Assumption

This section discusses certain characteristics of the disjunctive assumption from an empirical perspective.

1) BRB Input Set Is Complete: Suppose that input I has m dimensions (m attributes in BRB) and the mth attribute/dimension has an upper bound of ub_m and a lower bound of lb_m. Then, the completion of BRB requires the following:

$$\min(I_m) \geq \text{lb}_m \tag{6a}$$

$$\max(I_m) \geq \text{ub}_m. \tag{6b}$$

Because the rule activation mechanism does not change, the BRB input set under the disjunctive assumption is still complete.

We further illustrate this condition in Fig. 4.

Fig. 4(a) shows the conjunctive assumption, in which the adjacent rules (A_2B_2, A_2B_3, A_3B_3, and A_3B_2) are activated for input I. Fig. 4(b) shows the disjunctive assumption for a case in which the adjacent rules A_2B_2 and A_3B_3 are activated for input I. In fact, for any input under the disjunctive assumption, the adjacent rules could be activated using the rule activation procedure in [8] by projection onto the axes, which ensures

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

that the matching-degrees calculation results remain consistent with the conjunctive assumption.

2) BRB Is Downsized: Based on the conceptual diagram given in Fig. 4(b), the number of rules in BRB, K, under the disjunctive assumption is only related to the number of the referenced values of the attributes, as in (7). Thus, the size of BRB under the conjunctive assumption is determined by the number of attributes and the number of referenced values for those attributes, as in (2)

$$K = m_i. \tag{7}$$

3) Computational Requirements Are Reduced: Equations (8a/8b) calculate the number of parameters for BRB optimization under the conjunctive and disjunctive assumptions, respectively

$$\text{para}_{\text{con}} = n * \prod_{i=1}^{M} m_i + \prod_{i=1}^{M} m_i + \sum_{i=1}^{M} (m_i - 2) \tag{8a}$$

$$\text{para}_{\text{dis}} = n * K + K + \sum_{i=1}^{M} (K - 2) \tag{8b}$$

where para_{con} and para_{dis} denote the number of parameters under the conjunctive and disjunctive assumptions, respectively. m_i denotes the number of referenced values for the ith attribute, $i = 1, \ldots, M$, K denotes the number of rules in the BRB.

According to (8a/8b), para_{con} grows exponentially with an increasing number of attributes and/or attribute referenced values, while para_{dis} grows linearly. Therefore, the computational requirement for the disjunctive assumption is far reduced than the conjunctive assumption.

It also shows that the disjunctive assumption, which would be introduced in Section III, can significantly downsize BRB and reduce computational requirements.

Table I further shows an example of the number of (activated) rules in correlation with the number of attributes and the number of the referenced values of the attributes.

In Table I, the numerical example is assumed to have three attributes, and each attribute is assumed to have 2/3/4/5 referenced values. As is presented in Table I, the disjunctive assumption can help downsize BRB and reduce computational requirements.

TABLE I
NO. OF RULES AND PARAMETERS FOR BRB WITH THREE ATTRIBUTES UNDER EITHER ASSUMPTION

	conjunctive				disjunctive			
no. of ref. values	3	4	5	6	3	4	5	6
no. rules	27	64	125	216	3	4	5	6
no. activated rules	8	8	8	8	3	3	3	3
no. parameters	165	353	661	1308	21	30	39	48

IV. BRB JOINT OPTIMIZATION MODEL UNDER THE DISJUNCTIVE ASSUMPTION

A. New AIC-Based Objective for BRB Joint Optimization

For a given linear model

$$z(k) = h_1(k)\theta_1 + h_2(k)\theta_2 + \cdots + h_N(k)\theta_N + e(k) \tag{9}$$

where $z(k)$ is the output, $h_i(k)$ is the input, θ_i is the model parameter, and $e(k)$ is the model noise.

Akaike [1] proposed the following criterion to determine the smallest order (or number of independent parameters) for the model:

$$\text{AIC}(\hat{N}) = -2 \log L\left(\hat{\theta}_{\text{ML}}\right) + 2\hat{N} \tag{10}$$

where $\hat{\theta}_{\text{ML}}$ is the maximum likelihood estimate value of the parameter vector $\theta = [\theta_1, \theta_2, \ldots, \theta_N]$, $L(\hat{\theta}_{\text{ML}})$ is the likelihood function under $\hat{\theta}_{\text{ML}}$, and $\hat{N}$ is the number of independent parameters in the model.

Although AIC was developed for a linear model, it has been argued that AIC could also be applied to BRB for nonlinear models. Based on the deduction process given in [6], a new AIC-based objective (AIC for short) for BRB is given in (11), which is based on (10)

$$\text{AIC} = M \ln \sigma^2 + 2N \tag{11}$$

where σ^2 is the squared error between the actual outputs and estimated outputs, M is the number of the training data points, and N is the number of parameters in BRB.

σ^2 could be calculated using MSE with

$$\sigma^2 = M \bullet \text{MSE}. \tag{12}$$

Based on (11) and (12), we have

$$\text{AIC} = M \ln(M \bullet \text{MSE}) + 2N. \tag{13}$$

Equation (13) shows that when there are fewer parameters, modeling complexity is low, and the first part of the AIC contributes to a larger portion of its value. In this phase, AIC and MSE would both decrease with increasing numbers of parameters/rules. When the number of parameters/rules surpasses a certain number, MSE would still slowly decrease, but AIC would begin to increase, because the second part of AIC would contribute to a larger portion of its value.

Based on the above analysis, AIC can represent both modeling accuracy and complexity, which leads to the following section, in which a bi-level optimization model is constructed.

B. Bi-Level BRB Joint Optimization Model Under the Disjunctive Assumption

Based on the above analysis, the bi-level model could be derived as follows:

$$\begin{aligned} \min\ & \text{AIC}(\text{MSE}, k) \\ \min\ & \text{MSE}(v) \\ \text{s.t.}\ & v, k \end{aligned} \tag{14}$$

where v denotes a vector which is composed of multiple parameters to be optimized, including the referenced values of each attribute, x, the initial rule weight, θ, and the belief of each scale in each rule, β.

C. Upper-Level Optimization Model With AIC as the Objective

In the upper-level optimization model, the objective is to minimize AIC. The decisive variables are MSE, derived from lower-level optimization and the number of parameters.

The upper-level optimization model is as follows:

$$\min \ \mathrm{AIC} = f(\mathrm{MSE}, k) \tag{15a}$$
$$\text{s.t.} \ \ k \in [2, \ldots, K] \tag{15b}$$

where k is the number of rules. k should be higher than two, because there must be lower- and upper-level bounds for each attribute. K is the total number of rules.

D. Lower-Level Optimization Model With AIC as the Objective

For the lower-level optimization model, the objective is modeling accuracy, represented by MSE between the actual outputs and estimated outputs of the model. Since the number of rules is fixed, the decisive variables are as follows.

1) The referenced values for each attribute, A_i^k.
2) The initial rule weight, θ_k.
3) The belief of each scale in each rule, $\beta_{n,k}$.

For a given number of rules, k, the decisive variables form a vector, $\mathbf{v} = (A_i^k, \theta_k, \beta_{n,k})$.

The lower-level optimization model is given as follows:

$$\min \ \mathrm{MSE}(\mathbf{v}) \tag{16}$$

s.t.

$$\mathrm{lb}_i \leq A_i^k \leq \mathrm{ub}_i, \quad k = 1, \ldots, K; i = 1, \ldots, M \tag{17a}$$
$$A_i^1 = \mathrm{lb}_i \tag{17b}$$
$$A_i^K = \mathrm{ub}_i \tag{17c}$$
$$0 < \theta_k \leq 1 \tag{17d}$$
$$0 \leq \beta_{n,k} \leq 1, \quad n = 1, \ldots, N \tag{17e}$$
$$\sum_{n=1}^{N} \beta_{n,k} \leq 1 \tag{17f}$$

where (17a/17b/17c) denote that the referenced values of each attribute must fall within the upper and lower bounds, with the first and last referenced values of each attribute equal to the lower and upper bounds, respectively. Equation 17(d) denotes that the initial weight of each rule must be within (0, 1). Equations (17e/17f) sets the constraints on (the sum of) the beliefs.

V. BRB Joint Optimization Algorithm Under the Disjunctive Assumption

A. Upper-Level Optimization Algorithm With Local Optimization Strategy

For the upper-level optimization model in Section IV-C, the upper-level optimization algorithm is given as follows.

Step 1 (Initialization): Initialize certain parameters, including the initial number of rules (typically, $k_{\mathrm{ini}} = 2$), the number of attributes, and the number of scales in the conclusion part, etc.

Step 2: Go to the lower-level optimization with the initialized parameters.

Step 3: Calculate the value of AIC from (13), with the derived MSE from the lower-level optimization algorithm.

Step 4: Check AIC

```
if k > k_ini
    Compare AIC(k + 1) with AIC(k);
    if AIC(k + 1) < AIC(k)
        stop;
    end
else
    go to Step 6;
end.
```

Step 5: Check on the stop criterion

```
if k + 1 = K
    stop;
else
    go to Step 4;
end.
```

Step 6: Generate a new BRB with a local optimization strategy.

Step 6.1: Randomly select two neighboring rules, R_k and R_{k+1}, $(k > 1, k + 1 < K)$.

Step 6.2: Initialize one new rule, R'_k, between the two neighboring rules, R_k and R_{k+1}. Make sure that each parameter in R'_k satisfies the restrictions in (17a–17f).

Step 6.3: Local optimization for the new rule, R'_k. Considering the three rules, R_k, R'_k, and R_{k+1}, locally optimize the parameters of R'_k using only the related data (rather than the whole training dataset). The optimization model and algorithm are the same as that given in Sections IV-D and V-B.

$$\min \mathrm{MSE}\big(R_k, R'_k, R_{k+1}\big). \tag{18}$$

Step 6.4: Check for the stop criterion. The local optimization for R'_k stops either if it reaches a smaller MSE or the maximum number of generations

```
if MSE(R_k, R'_k, R_{k+1}) < MSE(R_k, R_{k+1})
    stop;
else
    if k = K
        stop;
    end
end.
```

Step 6.5: Generate the new BRB with the new added rules, R'_k.

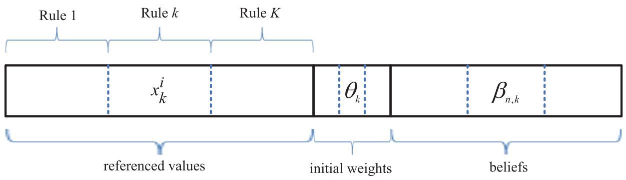


Fig. 5. Gene coding for the ith individual.

Remark 1: In the local optimization algorithm, the number of generations for the stop criterion does not need to be set very high; typically, 300 is an appropriate choice. Moreover, only the related original data is involved.

B. Lower-Level Optimization Algorithm Employing DE/GA/PSO As the Optimization Engine

The main steps of the lower-level optimization algorithm using ER and three evolutionary algorithms, namely genetic algorithm (GA), DE, and particle swarm optimization algorithm (PSO), as the inference and optimization engines are introduced as follows.

Step 1 (Parameter Initialization): For a given number of rules, n, the initiated parameters include the parameters for both BRB and the DE/GA/PSO algorithm. All parameters are initiated as individuals composed of decimal-coded genes and do not need to be decoded. For the ith individual, the genes are coded as shown in Fig. 5.

Step 2 (Operations): Conventional GA, DE, and PSO are applied as the optimization engine, respectively. Details of the three algorithms applied are listed as in [18], [21], and [27].

Step 3 (Fitness Function):

Step 3.1: (Rule Activation, Matching Degree Calculation, and Weight Calculation): The three procedures are conducted under the disjunctive assumption. Details can be found in [8].

Step 3.2 (ER Inference): After certain rules are activated, the K activated rules are integrated using ER, the analytic form of which is given in (19) and (20) and [34]

$$\beta_n = \frac{\mu\left[\prod_{k=1}^{K}\left(w_k\beta_{n,k} + 1 - w_k\sum_{n=1}^{N}\beta_{n,k}\right) - \prod_{k=1}^{K}\left(1 - w_k\sum_{n=1}^{N}\beta_{n,k}\right)\right]}{1-\mu\left[\prod_{k=1}^{K}(1-w_k)\right]} \tag{19}$$

$$\mu = \left[\sum_{n=1}^{N}\prod_{k=1}^{K}\left(w_k\beta_{n,k} + 1 - w_k\sum_{s=1}^{N}\beta_{s,k}\right) - (N-1)\prod_{k=1}^{K}\left(1 - w_k\sum_{s=1}^{N}\beta_{s,k}\right)\right]^{-1} \tag{20}$$

where β_n represents the belief for the nth degree.

Step 3.3 (Output by Utility): Suppose the utility of the nth scale D_n is denoted by $U(D_n)$. The expected utility of the belief distribution is given as

$$T = \sum_{n=1}^{N} U(D_n)\beta_n. \tag{21}$$

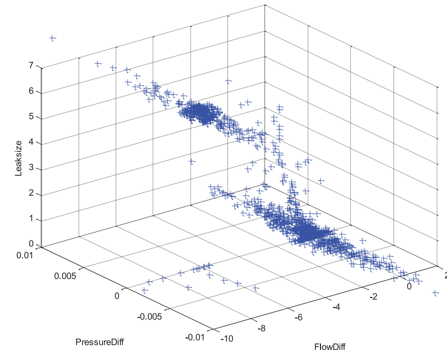


Fig. 6. Original data for the pipeline leak size detection case.

Step 4 (Selection): The ith individual, u_i^t, enters the new generation when its fitness function reaches a higher rated value, calculated by

$$x_i^{t+1} = \begin{cases} u_i^t & \text{if } f\left(u_i^t\right) \le f\left(x_i^t\right) \\ x_i^t & \text{otherwise} \end{cases} \tag{22}$$

where $f(\bullet)$ is the fitness function (MSE in this paper).

Step 5 (Check for the Stop Criterion): Check for whether the stop criterion (number of generations) is met. If not, go to step 3; if met, the individual with the smallest MSE is selected as the ultimate optimization result.

Remark 2: The reason that bi-level optimization is applied instead of a multiobjective optimization approach [2] is that AIC has been deduced to analytically represent the correlation between modeling accuracy and modeling complexity, and has thus become a single-objective optimization problem. Therefore, there is no need for a multioptimization-based approach. If a single objective was not obtained, then multiobjective optimization approaches should be applied.

Remark 3: A bi-level model is applied in order to reduce computational burden. In many practical applications, there may be over-numbered parameters to be optimized. By applying a bi-level model, the designed joint BRB optimization approach optimizes only the parameters with a given model structure, and then further optimizes the model structure. This is practical and more efficient from an engineering perspective.

VI. Case Studies

A. Pipeline Leak Detection Case

A pipeline leak detection case has been studied under multiple conditions [11], [38], [46]–[49]. Two factors influence the pipeline leak size: 1) pressure and 2) the flow difference between the two ends, namely *PressureDiff* and *Flowdiff*. The two factors are the BRB inputs.

A total of 2007 sets of data were gathered, as shown in Fig. 6, which indicates that the data are clearly influenced by noise.

Note: In Fig. 6, *PressureDiff* is [−0.01, 0.01] instead of [−0.02, 0.04] to better show the disparity of the data.

In previous studies, the total dataset was used as the testing dataset, while a different training dataset was selected.

TABLE II
COMPARISON RESULTS FOR CASE II OPTIMIZED BY DE/GA/PSO

optimization engine	number of rules identified					mis-identification
	3	4	5	6	others	(other than "3/4/5/6")
DE	10	6	4	5	5	16.67%
GA	13	1	3	6	7	23.33%
PSO	7	3	4	3	13	43.33%

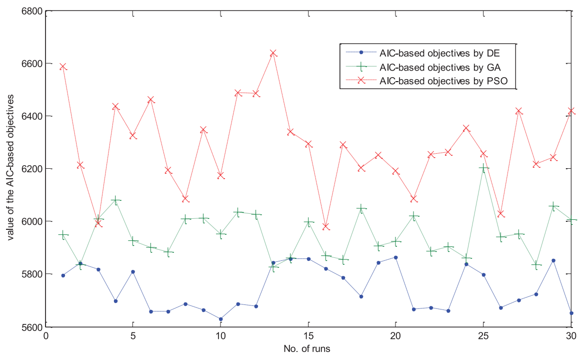


Fig. 7. AIC comparison for case II optimized by DE/GA/PSO.

References [38] and [11] chose 500 sets of data, [46] chose 900 sets of data, and [47] chose 800 sets of data. More importantly, those datasets were chosen from specific time intervals. In this paper, 500 sets of data were randomly selected as the training dataset.

Similar to previous studies, it is assumed that there are five scales in the conclusion part with the utilities as

$$\{D1, D2, D3, D4, D5\} = \{0, 2, 4, 6, 8\}. \tag{23}$$

Since the evolutionary algorithm is applied, there is no need to locate initial BRB. The settings are as follows: BRB is initiated with three rules, 40 individuals, and 2000 generations. The experiment is performed with 30 runs.

To study the efficiency of GA, DE, and PSO as the optimization engine, Table II presents the number of rules identified as the best fit over 30 runs of DE/GA/PSO.

As is presented in Table II, BRBs optimized by DE/GA/PSO all identity "three" as the optimal number of rules over 30 runs, although with different probability percentages. GA produces the highest percentage (43.33%, 13 out of 30 runs) while that for DE is 33.33% (10 out of 30 runs). Moreover, if we consider the probability of identifying "others" in Table II as "misidentifications," then PSO has the highest misidentification rates (43.33%, 13 out of 30 runs). Comparatively, the misidentification rates for GA and DE are rather low, at 17.67% and 23.33% (5 and 7 of 30 runs misidentified), respectively.

Fig. 7 further investigates the AICs for 30 runs derived by DE/GA/PSO. It shows that the AICs derived by PSO are far larger than those derived by GA, which are larger in turn than those derived by DE.

Further calculation results are presented in Table III, which indicates that the variance of AICs by DE over 30 runs was the smallest. Moreover, PSO identified ten rules as the optimal result, which is quite different from the optimal number of rules identified by DE and GA. In combination with the poor performance of AIC in Fig. 7, this indicates that PSO has failed in this optimization process.

TABLE III
AIC COMPARISON FOR THE PIPELINE CASE II
OPTIMIZED BY DE/GA/PSO

optimization engine	AIC			MSE	No. rules with the smallest AIC
	min	average	variance		
DE	5626.40	5673.21	3392.69	0.2917	5
GA	5826.01	5952.47	7818.03	0.3805	3
PSO	5980.94	6284.32	26669.25	0.7939	10

TABLE IV
BRB MODEL WITH FIVE RULES, OPTIMIZED BY DE

No.	weights	*FlowDiff*	*PressureDiff*	{D1, D2, D3, D4, D5}
1	0.9627	-10.0000	-0.2000	(0, 0.9998, 0, 0.0001, 0.0001)
2	0.0004	-0.5768	-0.0115	(0.9919, 0.0004, 0, 0.0030, 0.0047)
3	0.0041	-7.1136	-0.0130	(0.0493, 0.0908, 0.0423, 0.4389, 0.3787)
4	0.8387	-9.8702	-0.0143	(0.9997, 0, 0, 0.0002, 0.0001)
5	0.0000	2.0000	0.0400	(0.1420, 0.0238, 0.5193, 0.1604, 0.1545)

Note: The initial rule weights of rule No. 2/3/5 in Table IV were 3.9886E-04, 4.1172E-03, 2.0793E-08, respectively.

TABLE V
MSEs DERIVED IN PREVIOUS STUDIES

No.	refs	size (train)	size (test)	MSE (test)	no. rules	assum.
1	2007 [38]	500 (specific)	2008	0.4049	56	con.
2	2009 [47]	800 (specific)	2008	NA	56	con.
3	2010 [46]	900 (specific)	2008	0.7880	56	con.
4	2010 [48]	305 (specific)	17	0.0241	5	con.
5	2011 [11]	500 (specific)	2008	0.3990	56	con.
6	2016 [34]	900 (random)	2008	0.4450 (0.5040)	6(14)	con.
7	this study	500 (random)	2008	0.2917	5	dis.

B. Comparison Between Existing Studies and the BRB Joint Optimization Approach for Pipeline Leak Detection Case

Table IV shows a BRB with five rules, with the smallest AIC optimized by DE. The MSE of the BRB in Table IV is 0.2917, with an AIC of 5626.40.

Table V compares the results of different studies. It shows that the BRB achieved in this paper can produce results with higher accuracy and lower complexity (fewest rules) relative to previous studies. "con./dis." in the last column and "assum." in Table V stand for the conjunctive/disjunctive assumption, respectively.

Note that [48] derived a BRB with only five rules. However, this BRB was based on a specific segment of data (305 of the original 2008 sets of data were used as the training dataset, and 17 as were used the testing dataset) and the derived BRB was incomplete: certain inputs could not be handled (it was reasonable and feasible because in [48] the BRB is online updated and the testing dataset is very small).

Fig. 8 shows a comparison of the outputs and the errors for the BRBs in Table IV, derived by [46].

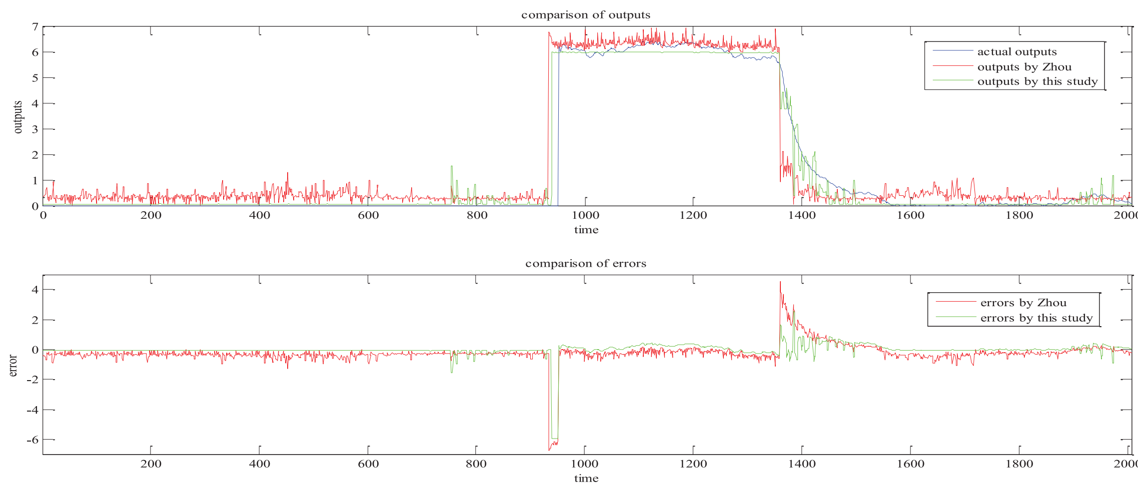


Fig. 8. Comparison of the outputs and errors for the pipeline leak detection case.

TABLE VI
COMPARISON ON MSE AMONG SVM/FUZZY SYSTEM/BRB FOR THE PIPELINE LEAK DETECTION CASE

	parameters settings			
fuzzy system	MF type	MFs: (3, 3)	MFs: (4,4)	MF: (5, 5)
	trimf	0.5073	0.6100	0.6859
	gbellmf	0.5971	0.6520	0.8737
	gauss2mf	0.6557	0.7884	0.5941
	pimf	0.5788	0.8716	0.7710
	dsigmf	0.5815	0.7716	0.5897
SVM	C	σ^2=1	σ^2=5	σ^2=10
	0.05	0.9144	0.9389	1.1976
	10	0.4219	0..4623	0.5232
	100	0.4269	0.4242	0.4439
	200	0.4669	0.4466	0.4291
BRB	0.2917			

C. Comparison Between SVM/Fuzzy System and BRB Joint Optimization Approaches for the Pipeline Leak Detection Case

To further demonstrate the efficiency of the proposed BRB joint optimization approach, results produced using SVM and an adaptive neural fuzzy system (ANFS) in multiple parameter settings are shown in Table VI.

For ANFS, "MFs: (3, 3)/(4, 4)/(5, 5)" stands for three/four/five parameters for each attribute (*FlowDiff* and *PressureDiff*) which produces 9/16/25 rules, MF type is linear.

For SVM, the two parameters are C and σ^2, ranging from 0.05 to 200, and 1 to 10, respectively.

As indicated by Table VI, the case study results for BRB outperform those for SVM and ANFS. With respect to SVM and ANFS, no conclusive comparison can be drawn, particularly given their different parameter settings.

D. Gas Turbine Engine Sensor Signal Inference Case

The second case aims to infer the signals from a gas turbine engine based on the signals from other working sensors.

Sensors are among the most crucial components in a gas turbine path. Accurate measurement and reliable work have a direct impact on the control, monitoring, and diagnosis systems of a turbofan engine [24]. Many sensors detect the pressure, temperature, and the rotor speed along the gas path of a turbofan engine. A fault occurring in any sensor could lead to incorrect engine control decisions or faulty diagnosis. Therefore, it is essential to reconstruct signals from the remaining functioning sensors. There are five sensors along the gas path, which can detect five parameters: 1) P_{t25} denotes the inlet pressure of high-pressure compressor; 2) T_{t3} denotes the outlet temperature of high-pressure compressor; 3) T_{t45} denotes the outlet temperature of high-pressure turbine; 4) N_h denotes the speed of high-pressure rotor; and 5) N_l denotes the speed of low-pressure rotor.

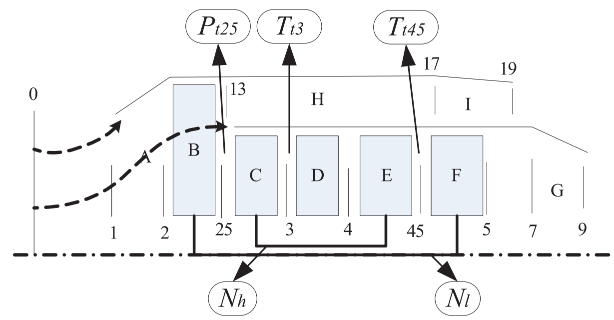


Fig. 9. Engine layout of the gas turbines case.

The engine layout of the gas turbine is shown in Fig. 9.

In an assumptive scenario, the data gathered from T_{t45} are inferred based on the data from the other four sensors (P_{t25}, T_{t3}, N_h, and N_l). There are 1000 sets of data gathered in total.

Present approaches to the gas turbine engine sensor signal inference problem or "out of the box" engine sensor diagnosis problem include model-based approaches and data-driven approaches. The main disadvantage of model-based approaches is that they operate under linear assumptions, which are impractical in many conditions [3]. Data-driven approaches have therefore become the prevailing choice, particularly those with an intelligent training and learning mechanism (SVM [30], neural networks [26], etc.)

TABLE VII
COMPARISON FOR CASE II OPTIMIZED BY DE/GA/PSO

optimization engine	number of rules identified				identification ("3/4/5" as the optimal result)
	3	4	5	others	
DE	3	18	9	0	100%
GA	0	7	5	18	40.00%
PSO	13	9	6	2	93.33%

TABLE VIII
AIC COMPARISON FOR CASE II OPTIMIZED BY DE/GA/PSO

optimization engine	AIC			MSE	No. rules
	min	average	variance		
DE	-1520.70	-1472.28	560.49	2.65E-05	4
GA	-1479.60	-1400.79	2170.99	2.71E-05	6
PSO	-1210.40	-983.86	13381.23	5.22E-05	5

The basic setup for using BRB as an inference model is as follows. The data gathered from P_{t25}, T_{t3}, N_h, and N_l are taken as the four attributes, while T_{t45} is taken as the inference result. The inference result for T_{t45} is assumed to be with five evenly distributed scales

$$\{D_1, D_2, D_3, D_4, D_5\} = \{0.8577, 0.9054, 0.9531, 1.0008, 1.0486\}. \quad (24)$$

Eight hundred sets of data are selected as the training dataset, while the total dataset is used as the testing dataset.

Because the evolutionary algorithm is applied, the initial solutions (BRB) are randomly generated. The other parameters are as follows: BRB is initiated from three rules, 20 individuals, and 500 generations. This experiment is performed with 30 runs.

To validate the efficiency of DE, GA, and PSO as the optimization engines, the numbers of rules identified as the best fit over 30 runs of DE/GA/PSO are compared. Table VII presents the optimal number of rules identified by DE/GA/PSO. Both DE and GA identify "4" as the optimal number of rules, while PSO identifies "3" as the optimal number of rules. If we adopt a broader perspective by considering 3, 4, and "5" as acceptable numbers of rules, then DE and PSO produce 100% and 93.33%, respectively, while GA produces only 40%. It can be concluded that GA produces less suggestive results than DE or PSO.

Fig. 10 further explores the values of the AICs; it shows that the AICs produced by PSO much larger than those produced by GA, which are larger in turn than those produced by DE. This indicates the failure of PSO in this optimization process. Closer observation of DE and GA shows that the AICs produced by DE are smaller than those produced by GA. A detailed AIC comparison of DE/GA/PSO optimizations, presented in Table VIII, also validates the conclusion that DE produced not only the optimal results (denoted by the smallest AICs), but also the most robust results (denoted by the smallest variance).

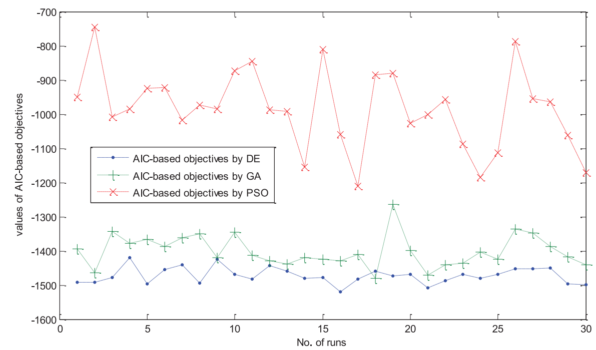


Fig. 10. AIC comparison for case II optimized by DE/GA/PSO.

TABLE IX
BRB MODEL WITH FOUR RULES AND ALSO OPTIMIZED BY DE

No.	weight	P_{t25}	T_{t3}	N_h	N_l	{D1, D2, D3, D4, D5}
1	0.3412	0.7063	0.8474	0.7757	0.8708	(0.9649, 0.0214, 0.0046, 0.0077, 0.0014)
2	0.1834	0.7755	0.9888	0.8281	1.0412	(0.4552, 0.1958, 0.0993, 0.2341, 0.0156)
3	0.0743	0.7369	1.0470	0.7949	1.0742	(0.4090, 0.1802, 0.2634, 0.1457, 0.0018)
4	0.4600	1.0815	1.0548	1.1093	1.0835	(0.0017, 0.0002, 0.0007, 0.0092, 0.9883)

TABLE X
COMPARATIVE RESULTS OF ANFS, SVM, AND BRB

	parameter settings			
ANFS	MF type	MFs: (3,3,3,3)	MFs: (4,4,4,4)	
	trimf	0.005189	0.005622	
	gbellmf	0.005331	0.003567	
	gauss2mf	0.005506	0.003437	
	pimf	0.005861	0.006886	
	dsigmf	0.005338	0.007758	
SVM	C	σ^2=0.01	σ^2=0.05	σ^2=0.1
	0.5	6.08E-04	1.17E-04	6.76E-05
	5	3.29E-04	7.70E-05	4.95E-05
	50	1.10E-03	2.01E-04	8.89E-05
BRB	2.65E-05			

E. Comparison Between Existing Studies and the Joint BRB Optimization Approach for the Gas Turbine Engine Case

The SVM approach is applied in a comparative fashion, because SVM has shown potential in pattern recognition [9], regression analysis [30], signal processing [19], etc. Multiple parameters settings are tested for SVM, and the results are presented in Tables X and XI. Of these settings, the optimal MSE using SVM ($C = 5, \sigma^2 = 0.1$) reaches 4.95E-05. Comparatively, the BRB with four rules optimized by DE in Table IX produces results with an MSE of 2.65E-05, which is 46.46% lower than that produced by SVM. This comparison is also shown in Fig. 11.

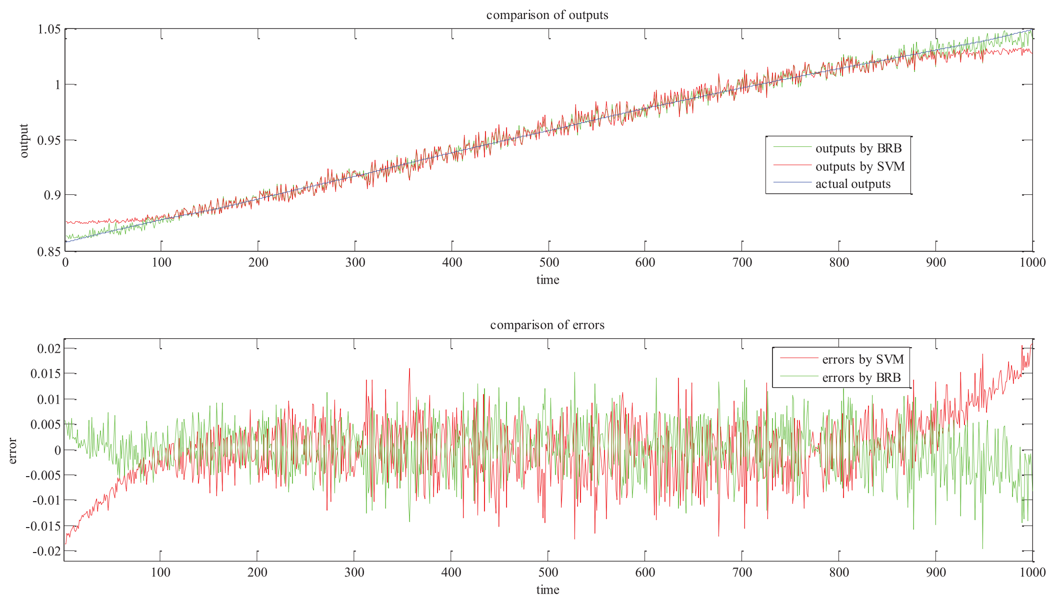


Fig. 11. Comparison of the results of SVM ($C = 10, \sigma^2 = 0.3$) and BRB for the gas turbine engine case.

TABLE XI
DETAILED COMPARISON BETWEEN SVM AND BRB

	1-200	201-800	801-1000
SVM	5.60E-05	3.52E-05	8.58E-05
BRB	1.17E-05	2.81E-05	3.65E-05

Table X presents a detailed comparison between the results derived by BRB, ANFS, and SVM for different parameter settings. As in Table X, BRB and SVM produce similar results, and both outperform ANFS, regardless of the parameter settings.

Table XI further compares the outputs and the errors for SVM and BRB models. It shows that there are, in fact, very limited differences between the results derived from SVM and BRB in the middle section. Most of the inaccurate modeling results in SVM occur at the two ends.

VII. CONCLUSION

This paper proposes a novel BRB structure and parameters joint optimization approach for complex nonlinear system modeling under the disjunctive assumption. Two challenges are addressed and solved. The first challenge is how to avoid the combinatorial explosion problem posed by the traditional conjunctive assumption during the construction of the BRB. The second challenge is how to jointly optimize both the structure and parameters of BRB.

For the first challenge, the conventional conjunctive assumption is replaced with a disjunctive assumption, to downsize the BRB while maintaining its completeness; this helps to avoid the combinatorial explosion problem and conserves computational effort. For the second challenge, a BRB joint optimization approach has been developed, combining a bi-level optimization model and optimization algorithm. In the joint optimization model, a new AIC-based objective is used to represent both modeling accuracy and modeling complexity. In the optimization algorithm, three evolutionary algorithms (DE, GA, and PSO), are comparatively tested as the optimization engine.

Two practical cases are studied to validate the efficiency of the BRB joint optimization approach. The case study results show that DE has superior performance in comparison with GA and PSO. Moreover, compared with the results of existing studies, the joint optimization approach effectively improves system modeling accuracy, as well as identifying the best decision structure for BRB with the smallest AIC. The proposed BRB joint optimization approach is further compared with ANFS and SVM for the two cases, further validating its efficiency.

Although the disjunctive assumption explored in this paper has produced superior performance for both case studies, it should be noted that there are limitations to using the disjunctive assumption in BRB construction. For example, in certain practical applications, only the conjunctive assumption should be applied when antecedent factors (BRB attributes) co-influence the results in a conjunctive fashion. In future studies, the correlations between the two assumptions should be explored.

REFERENCES

[1] H. Akaike, "Information theory and an extension of the maximum likelihood principle," in *Proc. 2nd Int. Symp. Inf. Theory, Tsahkadsor, Armenia, USSR*, Budapest, Hungary, Sep. 1971, pp. 267–281.

[2] R. Alcala, M. J. Gacto, and F. Herrera, "A fast and scalable multiobjective genetic fuzzy system for linguistic fuzzy modeling in high-dimensional regression problems," *IEEE Trans. Fuzzy Syst.*, vol. 19, no. 4, pp. 666–681, Aug. 2011

[3] R. Bettocchi and P. R. Spina, "ARX linear model set-up for fault diagnosis of gas turbine sensors," Amer. Soc. Mech. Eng., New York, NY, USA, Tech. Rep. 97-GT-027, 2007.

[4] L. L. Chang, M. J. Li, B. Cheng, and P. Zeng, "Integration-centric approach to system readiness assessment based on evidential reasoning," *J. Syst. Eng. Electron.*, vol. 23, no. 6, pp. 881–890, Dec. 2012.

[5] L. L. Chang, J. B. Sun, J. Jiang, and M. J. Li, "Parameter learning for the belief rule base system in the residual life probability prediction of metalized film capacitor," *Knowl. Based Syst.*, vol. 73, pp. 69–80, Jan. 2015.

[6] L. L. Chang, L. Y. Wang, W. Wang, G. Liu, and X. D. Ling, "Akaike information criterion-based objective for belief rule base optimization," in *Proc. 8th Int. Conf. Intell. Human Mach. Syst. Cybern.*, Zhejiang, China, 2016. pp. 545–549.

[7] L. L. Chang, Y. Zhou, J. Jiang, M. J. Li, and X. H. Zhang, "Structure learning for belief rule base expert system: A comparative study," *Knowl. Based Syst.*, vol. 39, pp. 159–172, Feb. 2013.

[8] L. L. Chang, Z. J. Zhou, Y. You, L. H. Yang, and Z. G. Zhou, "Belief rule based expert system for classification problems with new rule activation and weight calculation procedures," *Inf. Sci.*, vol. 336, no. 1, pp. 75–91, 2016.

[9] L. H. Chiang, M. E. Kotanchek, and A. K. Kordon, "Fault diagnosis based on Fisher discriminant analysis and support vector machines," *Compt. Chem. Eng.*, vol. 28, no. 8, pp. 1389–1401, 2004.

[10] Y.-W. Chen, J.-B. Yang, D.-L. Xu, and S.-L. Yang, "On the inference and approximation properties of belief rule based systems," *Inf. Sci.*, vol. 234, pp. 121–135, Jun. 2013.

[11] Y.-W. Chen, J.-B. Yang, D.-L. Xu, Z.-J. Zhou, and D.-W. Tang, "Inference analysis and adaptive training for belief rule based systems," *Expert Syst. Appl.*, vol. 38, no. 10, pp. 12845–12860, 2011.

[12] A. P. Dempster, "Upper and lower probabilities induced by a multivalued mapping," *Ann. Math. Stat.*, vol. 38, no. 2, pp. 325–339, 1967.

[13] W. R. Freudenburg, "Perceived risk, real risk: Social science and the art of probabilistic risk assessment," *Science*, vol. 242, no. 4875, pp. 44–49, 1988.

[14] A. Fuchs, *Nonlinear Dynamics in Complex Systems*. Heidelberg, Germany: Springer, 2013.

[15] H. Gargama and S. K. Chaturvedi, "Criticality assessment models for failure mode effects and criticality analysis using fuzzy logic," *IEEE Trans. Rel.*, vol. 60, no. 1, pp. 102–110, 2011.

[16] J. Gordon and E. H. Shorliffe, "A method for managing evidential reasoning in a hierarchical hypothesis space," *Artif. Intell.*, vol. 26, no. 3, pp. 323–357, 1985.

[17] N. R. Guo and T.-H. S. Li, "Construction of a neuron-fuzzy classification model based on feature-extraction approach," *Expert Syst. Appl.*, vol. 38, no. 1, pp. 682–691, 2011.

[18] J. H. Holland, *Adaptation in Natural and Artificial Systems*. Cambridge, MA, USA: MIT Press, 1975.

[19] L. B. Jack and A. K. Nandi, "Support vector machines for detection and characterization of rolling element bearing faults," *J. Mech. Eng. Sci.*, vol. 215, no. 9, pp. 1065–1074, 2001.

[20] J. Jiang, X. Li, Z.-J. Zhou, D.-L. Xu, and Y.-W. Chen, "Weapon system capability assessment under uncertainty based on the evidential reasoning approach," *Expert Syst. Appl.*, vol. 38, no. 11, pp. 13773–13784, 2011.

[21] J. Kennedy and R. Eberhart, "Particle swarm optimization," in *Proc. IEEE Int Conf. Neural Netw.*, Perth, WA, Australia, 1995, pp. 1942–1948.

[22] J. Lawrence, *Introduction to Neural Networks*. Nevada City, CA, USA: California Sci. Softw. Press, 1994.

[23] Z.-G. Liu, Q. Pan, J. Dezert, and G. Mercier, "Hybrid classification system for uncertain data," *IEEE Trans. Syst., Man, Cybern., Syst.*, to be published, doi: 10.1109/TSMC.2016.2622247.

[24] F. Lu, Y. Chen, J. Huang, D. Zhang, and N. Liu, "An integrated nonlinear model-based approach to gas turbine engine sensor fault diagnostics," *Proc. Inst. Mech. Eng. G J. Aerosp. Eng.*, vol. 228, no. 11, pp. 2007–2021, 2014.

[25] G. M. Mahmoud and E. E. Mahmoud, "Complex modified projective synchronization of two chaotic complex nonlinear systems," *Nonlin. Dyn.*, vol. 73, no. 4, pp. 2231–2240, 2013.

[26] T. Palme, M. Fast, and M. Thern, "Gas turbine sensor validation through classification with artificial neural networks," *Appl. Energy*, vol. 88, no. 11, pp. 3898–3904, 2011.

[27] K. V. Price, "Differential evolution vs. the functions of the 2nd ICEO," in *Proc. IEEE Int. Conf. Evol. Comput.*, Indianapolis, IN, USA, 1997, pp. 153–157.

[28] G. Rowe and G. Wright, "The Delphi technique: Past, present, and future prospects—Introduction to the special issue," *Technol. Forecast. Soc. Change*, vol. 78, no. 9, pp. 1487–1490, 2011.

[29] T. L. Saaty, *The Analytic Hierarchy Process*. New York, NY, USA: McGraw-Hill, 1980.

[30] B. Samanta, K. R. Al-Balushi, and S. A. Al-Araimi, "Artificial neural networks and support vector machines with genetic algorithm for bearing fault detection," *Eng. Appl. Artif. Intell.*, vol. 16, nos. 7–8, pp. 657–665, 2003.

[31] E.-E. Savan, J.-B. Yang, D.-L. Xu, and Y.-W. Chen, "A genetic algorithm search heuristic for belief rule-based model-structure validation," in *Proc. IEEE Int. Conf. Syst. Man Cybern.*, Manchester, U.K., 2013, pp. 1373–1378.

[32] L. W. Wagenhals, S. W. Liles, and A. H. Levis, "Toward executable architectures to support evaluation," in *Proc. Int. Symp. Collaborative Technol. Syst.*, Baltimore, MD, USA, 2009, pp. 502–511.

[33] Z. Wang, Y. Wang, and Y. Liu, "Global synchronization for discrete-time stochastic complex networks with randomly occurred nonlinearities and mixed time delays," *IEEE Trans. Neural Netw.*, vol. 21, no. 1, pp. 11–25, Jan. 2010.

[34] Y.-M. Wang, L.-H. Yang, Y.-G. Fu, L.-L. Chang, and K.-S. Chin, "Dynamic rule adjustment approach for optimizing belief rule-base expert system," *Knowl. Based Syst.*, vol. 96, pp. 40–60, Mar. 2016.

[35] Y.-M. Wang, J.-B. Yang, and D.-L. Xu, "Environmental impact assessment using the evidential reasoning approach," *Eur. J. Oper. Res.*, vol. 174, no. 3, pp. 1885–1913, 2006.

[36] X. Y. Wu and Q. Liu, *Weapon Test and Evaluation*. Beijing, China: Nat. Defense Ind. Press, 2008.

[37] X. J. Wu, M. S. Ma, Y. Q. Deng, and Z. H. Ma, "Navier–Stokes computations of a grid fin missile on hybrid structured–unstructured grids," *Acta Aerodyn. Sinica*, vol. 27, no. 4, pp. 419–423, 2009.

[38] D. L. Xu *et al.*, "Inference and learning methodology of belief-rule-based expert system for pipeline leak detection," *Expert Syst. Appl.*, vol. 32, no. 1, pp. 103–113, 2007.

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

[40] J.-B. Yang, J. Liu, D.-L. Xu, J. Wang, and H. Wang, "Optimization models for training belief-rule-based systems," *IEEE Trans. Syst., Man, Cybern. A, Syst., Humans*, vol. 37, no. 4, pp. 569–585, Jul. 2007.

[41] J.-B. Yang and D.-L. Xu, "Nonlinear information aggregation via evidential reasoning in multiattribute decision analysis under uncertainty," *IEEE Trans. Syst., Man, Cybern. A, Syst., Humans*, vol. 32, no. 3, pp. 376–393, May 2002.

[42] J.-B. Yang and D.-L. Xu, "Evidential reasoning rule for evidence combination," *Artif. Intell.*, vol. 205, pp. 1–29, Dec. 2013.

[43] L. A. Zadeh, "Fuzzy sets," *Inf. Control*, vol. 8, no. 3, pp. 338–353, 1965.

[44] X. A. Zhao, H. H. Chen, and M. He, "Simulation validation method for capability requirement of weapon system of systems based on Petri net," *J. Syst. Simulat.*, vol. 21, no. 4, pp. 1159–1163, 2009.

[45] Z.-J. Zhou, L.-L. Chang, C.-H. Hu, X.-X. Han, and Z.-G. Zhou, "A new BRB–ER-based model for assessing the lives of products using both failure data and expert knowledge," *IEEE Trans. Syst., Man, Cybern. Syst.*, vol. 46, no. 11, pp. 1529–1543, Nov. 2016.

[46] Z.-J. Zhou *et al.*, "A sequential learning algorithm for online constructing belief-rule-based systems," *Expert Syst. Appl.*, vol. 37, no. 2, pp. 1790–1799, 2010.

[47] Z.-J. Zhou, C.-H. Hu, J.-B. Yang, D.-L. Xu, and D.-H. Zhou, "Online updating belief rule based system for pipeline leak detection under expert intervention," *Expert Syst. Appl.*, vol. 36, no. 4, pp. 7700–7709, 2009.

[48] Z.-J. Zhou, C.-H. Hu, D.-L. Xu, J.-B. Yang, and D.-H. Zhou, "New model for system behavior prediction based on belief rule based systems," *Inf. Sci.*, vol. 180, no. 24, pp. 4834–4864, 2010.

[49] Z.-J. Zhou, C.-H. Hu, J.-B. Yang, D.-L. Xu, and D.-H. Zhou, "Online updating belief-rule-base using the RIMER approach," *IEEE Trans. Syst., Man, Cybern. A, Syst., Humans*, vol. 41, no. 6, pp. 1225–1243, Nov. 2011.

Lei-Lei Chang received the B.Eng. degree from Central South University, Changsha, China, in 2008, and the M. Eng. and Ph.D. degrees from the National University of Defense Technology, Changsha, in 2010 and 2014, respectively.

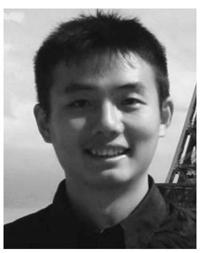
He is currently a Lecturer with the High-Tech Institute of Xi'an, Xi'an, China. He has published one book and approximately 20 articles. His current research interests include BRB structure and parameter learning and complex system modeling.

Zhi-Jie Zhou received the B.Eng. and M.Eng. degrees from the High-Tech Institute of Xi'an, Xi'an, China, in 2001 and 2004, respectively, and the Ph.D. degree from Tsinghua University, Beijing, China, in 2010.

He is an Associate Professor with the High-Tech Institute of Xi'an. He has published approximately 70 articles. His current research interests include belief rule base, dynamic system modeling, and hybrid quantitative and qualitative decision modeling.

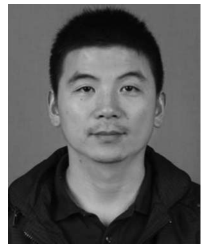
Yu-Wang Chen received the Ph.D. degree in control theory and control engineering from Shanghai Jiao Tong University, Shanghai, China, in 2008.

He is currently a Senior Lecturer of Decision Sciences with the Alliance Manchester Business School, University of Manchester, Manchester, U.K. His current research interests include decision and risk analysis under uncertainties, modeling, and optimization of complex systems, operational research, and data analytics.



Tian-Jun Liao received the bachelor's degree from the National University of Defense Technology, Changsha, China, in 2007, and the Ph.D. degree from the Institute de Recherches Interdisciplinaires et de Developpements en Intelligence Artificielle, Universite Libre de Bruxelles, Brussels, Belgium, in 2013.

He is currently an Assistant Researcher with the Beijing Institute of System Engineering, Beijing, China. His current research interests include heuristic optimization algorithms and automated algorithm configuration.



Yu Hu received the B.Eng., M.Eng., and Ph.D. degrees from the High-Tech Institute of Xi'an, Xi'an, China, in 2007, 2009, and 2014, respectively.

He is currently a Lecturer with the High-Tech Institute of Xi'an. He has published approximately 12 articles. His current research interests include turbine engine modeling, gas path fault diagnosis and sparse theory, and expert-systems-based methods and approaches.

Long-Hao Yang received the B.Eng. degree from Fuzhou University, Fuzhou, China, in 2012, where he is currently pursuing the Ph.D. degree.

He has published approximately 10 articles in BRB-related studies. His current research interests include BRB training and learning, machine leaning and artificial intelligence, and BRB applications in social economy and complex systems modeling.