Knowledge-Based Systems 142 (2018) 220–240

Contents lists available at ScienceDirect

Knowledge-Based Systems

journal homepage: www.elsevier.com/locate/knosys

A joint optimization method on parameter and structure for belief-rule-based systems

Long-Hao Yang[a,b], Ying-Ming Wang[a,*], Jun Liu[c], Luis Martínez[b]

[a] Decision Sciences Institute, Fuzhou University, Fuzhou, Fujian, PR China
[b] Department of Computer Science, University of Jaén, Jaén, Spain
[c] School of Computing and Mathematics, Ulster University, Northern Ireland, UK

ARTICLE INFO

Article history:
Received 5 April 2017
Revised 1 October 2017
Accepted 29 November 2017
Available online 2 December 2017

Keywords:
Belief-rule-based system
Parameter optimization
Structure optimization
Generalization error
Heuristic strategy

ABSTRACT

The belief-rule-based system (BRBS) is one of the most visible and fastest growing branches of decision support systems. As the knowledge base in the BRBS, the belief-rule-base (BRB) is required to be equipped with the optimal parameters and structure, which means the optimal value and number of parameters, respectively. Several optimization methods were therefore proposed in the past decade. However, these methods presented different limitations, such as the use of the incomplete parameter optimization model, lack of structure optimization, and so on. Moreover, it is impracticable to determine the optimal parameters and structure of a BRB using the training error because of over-fitting. The present work is focused on the joint optimization on parameter and structure for the BRB. Firstly, a simple example is utilized to illustrate and analyze the generalization capability of the BRBS under different numbers of rules, which unveils the underlying information that the BRBS with a small training error may not have superior approximation performances. Furthermore, by using the Hoeffding inequality theorem in probability theory, it is a constructive proof that the generalization error could be a better choice of criterion and measurement to determine the optimal parameters and structure of a BRB. Based on the above results, a heuristic strategy to optimize the structure of the BRB is proposed, which is followed by a parameter optimization method using the differential evolution (DE) algorithm. Finally, a joint optimization method is introduced to optimize the parameters and structure of the BRB simultaneously. In order to verify the generality and effectiveness of the proposed method, two practical case studies, namely oil pipeline leak detection and bridge risk assessment, are examined to demonstrate how the proposed method can be implemented in the BRB under disjunctive and conjunctive assumptions along with their performance comparative analysis.

1. Introduction

The decision support system is an important tool to address decision problems and thus many kinds of decision support systems have been constructed to deal with the product customization [1–3], medical diagnosis [4,5], and others. Among these decision support systems, the rule-based system has become one of the most useful approaches to model and analyze decision problems using various types of knowledge [6]. Traditional rule-based systems usually adopt simple IF-THEN rules to represent knowledge, e.g. in the form of "*IF presence of creatinine THEN renal failure is definite*" [7]. The rule means that the consequent "*renal failure is definite*" is believed to be true with the probability of 100% from the given cause "*presence of creatinine*". However, previous studies [8–10] have shown that such strict knowledge representation scheme is inefficient in expressing information with uncertainty, such as how to represent the consequent with the probability of 10%, 20%, or 50%. Therefore, there is a need to enhance the knowledge representation scheme in rule-based systems.

Due to the limitations of the traditional IF-THEN rules, many attempts [11,12] have been made to combine other theories, such as the interval probability theory and fuzzy set theory, to improve the capability of expressing uncertain information using IF-THEN rules. The prevailing one is the belief rule, which is a more advanced IF-THEN rule proposed by Yang et al. [13]. One of distinctive improvements is to replace the single-valued consequent by using a distributed assessment called a belief structure. The use of belief structure has enabled various kinds of uncertain information, including probability, fuzzy, and incompleteness. By grouping belief rules into one belief-rule-base (BRB), the belief-rule-based system (BRBS) has an effective knowledge base to address different deci-

* Corresponding author.
E-mail address: ymwang@fzu.edu.cn (Y.-M. Wang).

https://doi.org/10.1016/j.knosys.2017.11.039
0950-7051/

sion problems. To date, the applications of the BRBS include different areas, such as the prediction of consumer preferences [14–16], detection of pipeline leak [17,18], prediction of Forex trading [19], Assessment of product lives [20], safe evaluation of engineering systems [21–23], clinical decision support [24–26] and basic classification problems [27–29].

However, two basic issues for setting up a BRB should be solved before the corresponding BRBS is ready for an application.

(1) Determination of the optimal parameters involved in the BRB, which means to optimize the value of parameters for a BRB, such as the antecedent attribute weights, the rule weights, the utility values of the referential values used for each antecedent attribute, and the belief degrees attached to the consequent. The optimal parameters are closely related to the performance of the established BRBS. Accordingly, many parameter optimization models and techniques have been developed, such as the local parameter optimization model [30], the adaptive parameter optimization model [31], and the swarm intelligent algorithm-based optimization technique [26,32], etc.
(2) Determination of the optimal structure involved in the BRB, which means to optimize the number of parameters for a BRB. The optimal structure is closely linked to the size of BRB and affects the performance of the BRBS in terms of both accuracy and efficiency. The big number may lead to an over-fitting, whilst too small number may result in an under-fitting. Thus, many structure optimization approaches, including the sequential learning algorithm [31], dimensionality reduction-based structure learning method [33], the dynamic rule adjustment approach [34], etc, have been proposed in order to obtain an optimal structure of the BRB.

Based on the literature review for solving the above two basic issues, several deficiencies could be found. On the one hand, some attempts may not be able to produce promising results in optimizing a BRB because of their limitations. For instance, the local parameter optimization model [30] cannot achieve better accuracy than the adaptive parameter optimization model [31] under the same experimental conditions. On the other hand, some attempts are only available to determine the optimal parameters [30,31] without the optimal structure of a BRB. It is evident that both are important and necessary for constructing an optimal BRB [35]. More importantly, for most of attempts to optimize the parameters and the structure of a BRB, such as the sequential learning algorithm [35] and the dynamic rule adjustment approach [34], the minimal training error obtained by optimizing the parameters is regarded as an important criterion to determine the optimal structure. However, the minimal training error might be derived from the BRBS with the over-fitting.

To overcome the above limitations, the following challenges need to be addressed:

(1) Identifying a new criterion to determine the optimal parameters and structure of a BRB to avoid the over-fitting.
(2) Designing an efficient strategy to obtain the optimal one from the large number of possible the structure of BRBs.
(3) Developing a generic optimization technique to obtain the optimal parameters of a BRB from arbitrary a structure.

In the present work, to address the first challenge, a simple example is used to analyze and illustrate relationship between the generalization capability of BRBS and the size of BRB. Thus, three specific BRBs with different numbers of rules are designed to reveal the relationship between the approximation error and the desired number of rules. Furthermore, by using the Hoeffding inequality theorem [36] to link the training error, the number of training data, and the structure of the BRB, the concept of generalization error is proposed as a new criterion to determine the optimal parameters and structure of a BRB.

For the second challenge, the large number of possible BRBs options force that an efficient strategy is required to be proposed to obtain the optimal structure of a BRB. For example, based on the arbitrary number of referential values that can construct different BRBs with different structures, the optimal structure of the BRB has to be obtained from $2^6=64$, $2^8=256$ and $2^{10}=1024$ possible BRBs options if with two antecedent attributes and each attribute used three, four and five referential values respectively. In the present work, the set of referential values is used as a concise scheme to express the structure of a BRB. Then a heuristic strategy based on the generalization error is proposed to efficiently optimize the number of referential values for each antecedent attribute among all the possible BRBs.

For the third challenge, although different parameter optimization techniques of the BRB have been developed, three dilemmas must be addressed: (1) the stop criterion has a great impact on optimization results; (2) different parameter optimization models have different constraints for the same parameter, some of which may be dynamic; (3) the referential values are normally expressed as linguistic terms, so there should be an equivalent transformation between the utility values of the referential values and the corresponding linguistic terms. Accordingly, in the present work, the differential evolutionary (DE) algorithm [37] is adopted as a generic technique to address the above dilemmas in order to obtain the optimal parameters of a BRB under arbitrary a structure.

On the basis of the proposed solutions for the above three challenges, a new joint optimization method is proposed to obtain both the optimal parameters and the optimal structure of the BRB simultaneously. The key idea is summarized as follows: the above mentioned heuristic strategy and the DE algorithm-based technique are responsible for optimizing the structure and the parameters of the BRB, respectively. The generalization error is then used to judge whether the BRB is superior to other BRBs in terms of the structure and parameters. Finally, the BRB with the minimal generalization error is selected as the optimal one.

To illustrate the generality and effectiveness of the proposed method, the widely used case study on oil pipeline leak detection [17,18] and bridge risk assessment [38,39] are conducted on the BRB under conjunctive and disjunctive assumptions. Two main aspects, namely, accuracy and the number of parameters, are used to compare with other BRBSs and the corresponding parameter and structure optimization methods.

The rest of the paper is organized as follows: Section 2 briefly reviews the basics of the BRBS and its related works. Section 3 introduces and illustrates the concept of generalization error. Section 4 proposes the joint optimization method for obtaining the optimal parameters and the optimal structure of a BRB. Section 5 provides two case studies to demonstrate the generality and effectiveness of the proposed method, and the paper is concluded in Section 6.

2. Brief introduction to BRBS

2.1. Basics of BRBS

A BRBS consists of two main components. The first is the BRB that can be regarded as the knowledge base to save various kinds of uncertain information. The second is the evidential reasoning (ER)-based inference method that provides an inference engine to reply input data according to the BRB. A simple methodological framework of the BRBS is shown in Fig. 1.

For a BRB, suppose it has M antecedent attributes and one consequent attribute, each antecedent attribute U_i ($i=1,\ldots,M$) has J_i referential values $A_{i,j}$ which are used for describing the ith an-

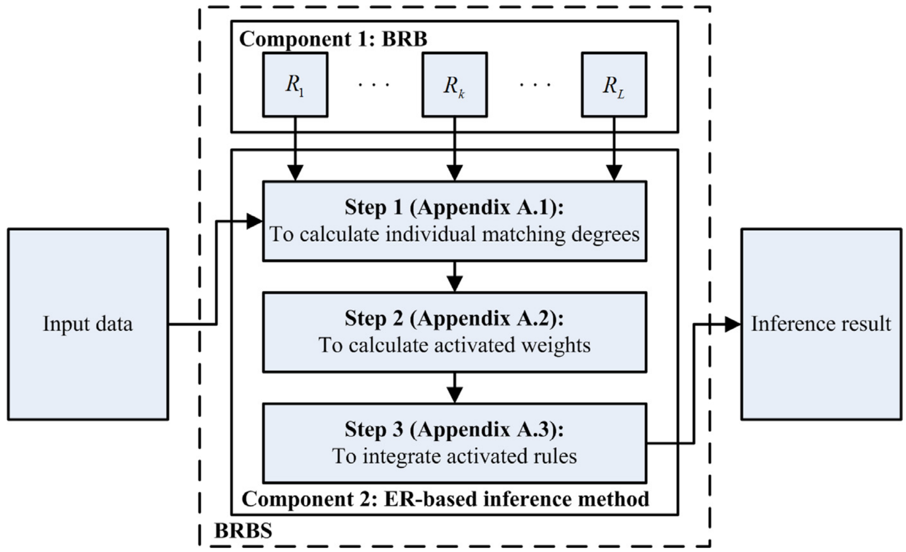

Fig. 1. Methodological framework of BRBS.

tecedent attribute U_i (in most of cases $A_{i,j}$ is expressed as the linguistic terms or the categorical terms) and the consequent attribute D has N referential values D_n ($n=1,...,N$) which are used for describing the consequent attribute D (in most of cases D_n also is expressed as the linguistic terms or the categorical terms). Thus, a belief rule from the BRB can be written as [13]:

$$\begin{aligned} &R_k : IF\ \left(U_1\ is\ A_1^k\right)\ AND/OR \cdots AND/OR\ \left(U_M\ is\ A_M^k\right), \\ &\quad THEN\ D\ is\ \{(D_n, \beta_{n,k});\ n = 1, \ldots, N\}, \\ &\quad with\ rule\ weight\ \theta_k\ and\ antecedent\ \ attribute\ , \\ &\quad weights\ \{\delta_1, \ldots, \delta_M\} \end{aligned} \tag{1}$$

where $\beta_{n,\,k}(0 \le \beta_{n,\,k} \le 1)$ denotes the belief degree to the referential value D_n in the kth ($k=1,...,L$) rule with $\sum_{n=1}^{N} \beta_n^k \le 1$; $A_i^k \in \{A_{i,j};\ j=1,...,J_i\}$ ($i=1,...,M$); δ_i denotes the weight of the ith antecedent attribute; θ_k denotes the weight of the kth rule; "*AND*" and "*OR*" denote the logical connectives that are to link different antecedent attributes with conjunctive and disjunctive assumptions, respectively.

Based on the above BRB, the ER-based inference method is used to integrate the belief rules with the activated weights associated with input data. As shown in Fig. 1, it is composed of three steps: (1) calculation of individual matching degrees in response to the input data to each antecedent attribute; (2) calculation of activated weights for each belief rule; and (3) integration of activated rules using the ER algorithm. The process of calculating activated weights for each rule is classified into two cases, therefore the cumulative and the multiplicative formulas are used respectively because the BRB may be constructed under the disjunctive or conjunctive assumptions. The detailed description of those steps can be found in Appendix A.

2.2. Construction and optimization of BRB

The construction and optimization of a BRB are crucial processes to improve the performance of the BRBS. Fig. 2 describes these processes for the BRB.

In the BRB construction process, expert knowledge provided by the domain experts are the main approach to establish belief rules, i.e., the domain experts have to determine which antecedent and consequent attributes need to be used, how many referential values should be used, and what kind of fuzzy membership functions or numerical values should be assigned for each referential value. Afterwards, by using logical connectives "*AND*" and "*OR*" to link different antecedent attributes, and providing initialized belief distribution of the consequent attribute, the initialized belief rules, namely initial BRB, can be constructed for a BRBS.

However, the initial BRB constructed by the expert knowledge may not lead to a desired accuracy because the domain experts are not always able to accurately determine the parameters and structure involved in the BRB due to lack of information, knowledge, and data. Hence, the parameter and structure optimization have to be done to improve the performance of the BRBS.

In the BRB optimization process, the same input data are provided for the actual system and the BRBS. The inference output of the BRBS may usually be different from the output of the actual system. Obviously, it is desirable that the difference between two outputs keep as small as possible. Based upon this viewpoint, the parameters and structure of the BRB must be trained to reduce these differences. Finally, the trained parameters and structure are used to update the initial BRB.

2.3. Related works on the parameter and the structure optimization

For the parameters and structure optimization of the BRB, it has attracted much attention so that many attempts have been made to obtain optimal parameters and structure of the BRB in the past decade.

Yang et al. [30] firstly proposed a parameter optimization model in order to obtain the optimal parameters of a BRB and it can be solved as a nonlinear constrained optimization problem shown as follows:

$$Minimize\ \sum_{t=1}^{T} {\xi_t}^2 = \sum_{t=1}^{T} \left(f(\boldsymbol{x_t}) - y_t\right)^2, \tag{2a}$$

$$Subject\ to\ \sum_{n=1}^{N} \beta_{n,k} = 1,\ k = 1, \ldots, L, \tag{2b}$$

$$0 \le \beta_{n,k} \le 1,\quad n = 1, \ldots, N;\ k = 1, \ldots, L, \tag{2c}$$

$$0 \le \theta_k \le 1,\ k = 1, \ldots, L, \tag{2d}$$

$$0 \le \delta_i \le 1,\ i = 1, \ldots, M, \tag{2e}$$

$$u(D_i) < u(D_j)\ if\ i < j,\ i, j = 1, \ldots, N, \tag{2f}$$

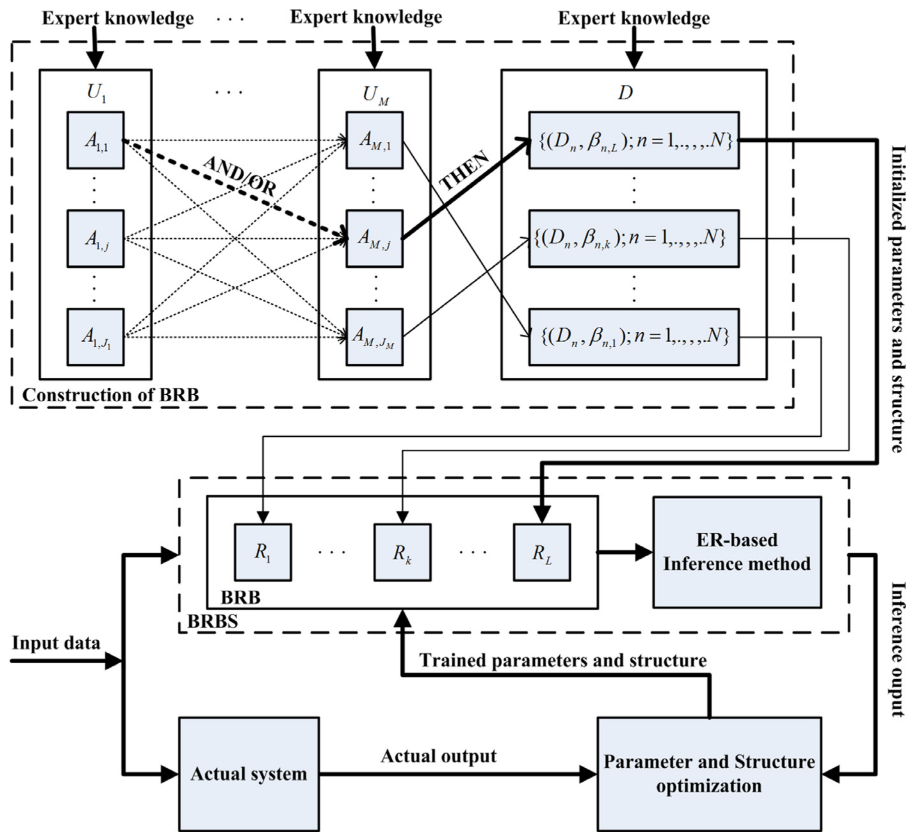

Fig. 2. Construction and optimization of BRB.

where $f(\boldsymbol{x_t})$ denotes the inference output of a BRBS given the input data $\boldsymbol{x_t}$ (See Appendix A for details), here $\boldsymbol{x_t}=(x_{1t}, x_{2t}, \ldots, x_{Mt})$, M is the number of antecedent attributes involved in the BRB, y_t is the actual output of the tth input data; Eqs. (2b)–(2c) are constraints on the belief degree; Eqs. (2d)–(2e) are on the antecedent attribute weights and the rule weights, respectively; and Eq. (2f) is the constraint on the utility values of the referential values used for consequent attribute. This model is implemented using the FMINCON function (a nonlinearity constrained optimization solver) in the MATLAB optimization toolbox. However, the implementation is falling easily into the local optimal solution and has very slow convergence speed.

For a further improvement, Chen et al. [31] constructed an adaptive parameter optimization model by adding the utility values of the referential values used for each antecedent attribute (with the constraint shown in Eq. (3)) as additional parameters to be trained into the model in [30], it was proved to obtain better performance of BRBS with the global training. The deficiencies of this model however are that it is still implemented using the FMINCON function in the MATLAB optimization toolbox and the utility values of the referential values used for the consequent attribute was not considered as the parameters to be trained.

$$u(A_{i,j}) - u(A_{i,j+1}) \le V_i,\ i = 1, \ldots, M;\ j = 1, \ldots, J_i - 1, \tag{3}$$

where Eq. (3) is the constraint on the utility values of the referential value in the ith antecedent attribute; V_i denotes the difference between the utility values of two adjacent referential values used for the ith antecedent attribute.

Zhou et al. [18] took into account expert knowledge for constructing a new parameter optimization model with expert intervention. The model can optimize the parameters of a BRB not only based on the input-output data pairs but also on the subjective expert knowledge. However, it is a pity that the expert intervention involves the complicated calculation process, which is difficult to understand and not easy to be implemented.

Chang et al. [40] studied the rule activation process of the BRBS and proposed a new parameter optimization model for the BRB under disjunctive assumption. Comparing to the model under conjunctive assumption, the following constraints were added into the model in [30].

$$lb_i \le u(A_i^k) \le ub_i,\ i = 1, \ldots, M;\ k = 1, \ldots, L, \tag{4a}$$

$$u(A_i^1) = lb_i, i = 1, \ldots, M, \tag{4b}$$

$$u(A_i^L) = ub_i, i = 1, \ldots, M, \tag{4c}$$

where Eqs. (4a)–(4c) are constraints on the utility values of the referential values regarding the antecedent attribute in the kth rule, the 1st rule, and the Lth rule, respectively; lb_i and ub_i denote the lower and upper bounds of the utility values of the referential values for the ith antecedent attribute.

However, the above constraints shown in Eqs. (4b)–(4c) usually lead to the local optimal parameters of a BRB because of the strict constraints for the utility values of the referential values in the 1st rule and the Lth rule. Thus, Yang et al. [41] further investigated the parameter optimization model in [40] and discussed the reason for the local optimal parameters of the BRB under disjunctive assumption. A new parameter optimization model was then proposed and the highest/lowest utility values of the referential values regarding the antecedent attribute are required to meet the following dynamic constraints in the specific rules.

$$u(A_i^s) = lb_i, s = \arg \min_{k=1,\ldots L} \{u(A_i^k)\}; i = 1, \ldots, M, \tag{5a}$$

$$u(A_i^s) = ub_i, s = \arg \max_{k=1,\ldots L} \{u(A_i^k)\}; i = 1, \ldots, M. \tag{5b}$$

For improving optimization techniques, Zhou et al. [42] analyzed the real-time behavior of the MATLAB-based optimization techniques and concluded that they are not suitable for optimizing the parameters of a BRB in a dynamic fashion. Therefore, an online optimization technique based on the expectation maximization (EM) algorithm was proposed to achieve the parameter optimization of the BRB. However, the proposed technique is essentially an iterative algorithm, and it failed to prove the efficiency due to lack of the comparative experimental studies with other techniques.

Zhou et al. [25] further discussed the MATLAB-based optimization techniques and stated that those techniques are sensitive to the initialization and are limited by their weak local searching abilities. Thus, an efficient global optimization technique based on the clonal selection algorithm (CSA) was proposed to optimize the parameters of BRB. Some similar optimization techniques based on other swarm intelligent algorithms were proposed in [26] and [32].

The above mentioned researches, however, all assumed that the structure of a BRB has been given in advance based on the expert knowledge. In most cases, the expert knowledge is not always available to determine the structure of a BRB so that it is necessary to consider obtaining both the optimal parameters and the optimal structure of the BRB.

Wang et al. [14] utilized the Principle Component Analysis (PCA) to downsize number of antecedent attributes in a BRB before the parameter optimization model and technique were used to optimize the parameters of the BRB. The achievement has been successfully applied in the consumer preference prediction. The similar methods were proposed in [15] and [16].

Chang et al. [33] compared several methods including the Grey Target, Multidimensional Scaling, Isomap, and PCA in downsizing the size of the BRB and the results shown that the PCA is the best one. However, these attempts based on the dimensionality reduction show the poor correlation between parameter optimization and structure optimization.

Zhou et al. [35] introduced the concept of statistical utility to automatically prune or add rules according to the desired training error. However, the most significant deficiency is that the proposed approach failed to ensure the completeness of a BRB. The incomplete BRB could not activate any rule when some special input data are given for the BRBS.

Wang et al. [34] used the density analysis to automatically reduce rules when the rule is closer to other rules, and then used the error analysis to automatically add a new rule when the training error of the BRBS is bigger than the desired threshold. As a result, the best decision structure of the BRB can be constructed from the incomplete or over-complete structure of the BRB.

Recently, Chang et al. [43] proposed a bi-level optimization method with Akaike Information Criterion (AIC)-based objective to achieve the parameter and structure optimization of the BRB simultaneously. It was noted that the enumeration strategy was introduced to determine how many rules should be kept in a BRB, which is time-consuming for constructing the compact structure of a BRB.

The above literature review clearly shows that the studies in both parameter and structure optimization can make the application of the BRBS more valuable in practices, but most attempts in this direction are still far from satisfaction. Furthermore, except for the method in [43], all other attempts to achieve structure optimization are based on the training error, which may be impracticable based on the optimization analysis in Section 3 below. Although the method in [43] takes the AIC to evaluate the optimal structure of the BRB, it requires considering all the possible rules in the BRB. This is also not easy when there are hundreds or thousands of rules to be considered. Therefore, a new method to achieve parameter optimization and structure optimization simultaneously is the final goal of the present work based on some preliminary studies.

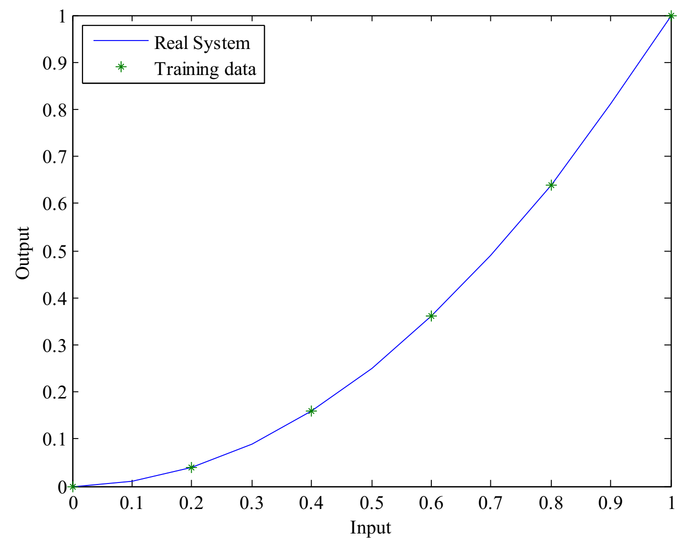

Fig. 3. Six training data obtained from an increasing function.

3. Analysis of generalization capability of BRBS

The generalization capability is an important criterion for various prediction models [44–46], which can be identified as the ability of the prediction model to have persistent performances on the unseen data. Thus, it is necessary to analyze the generalization capability of the BRBS and then to provide a better guide for parameter and structure optimization.

3.1. Generalization capability of BRBS under different number of rules

In the previous study [47], the BRBS has been proved to be a universal approximator so that it can approximate any real continuous function on a compact set with arbitrary accuracy. Thus, the generalization capability of the BRBS is described that the approximation error can be arbitrarily small while there are a large number of belief rules. However, the collected training data have direct influences on the accuracy of the BRBS results. According to the structure partition for the BRB [34], there are three kinds of BRB structures based on the number of rules under a given number of training data. If the number of rules for a BRB is less than the lower-bound threshold, the accuracy of the BRBS will decrease by decreasing number of rules. Correspondingly, if the number of rules is greater than the upper-bound threshold, the accuracy of the BRBS will decrease by increasing number of rules. Hence, the BRBS may not always have the capability of approximating continuous function. The example below is used to further illustrate the generalization capability of the BRBS under different numbers of rules.

Example 1. Suppose there is a single input system shown in Fig. 3, whose input is bounded in the interval [0, 1] and whose output $g(x)=x^2$ is a monotonically increasing function of the input x. To construct the BRBS, the following information is required:

(1) A set of training data, for example, a set of six input-output data pairs as follows:

$$\left\{(x_i, x_i^2), i=1,\ldots,6\right\} = \{(0,0), (0.2,0.04), (0.4,0.16), (0.6,0.36), (0.8,0.64), (1,1)\}. \quad (6)$$

Table 1
Three trained BRBs under different numbers of rules.

Type	Attribute weight	Rule no.	Rule weights	x	Belief degrees
BRBS-1	1.0000	1	1.0000	–	{0.0000, 0.2667, 0.7333, 0.0000, 0.0000}
BRBS-2	0.6314	1	1.0000	0.0000	{0.0080, 0.9893, 0.0000, 0.0000, 0.0027}
		2	0.5806	1.0000	{0.0000, 0.0628, 0.0000, 0.8116, 0.1256}
BRBS-6	1.0000	1	1.0000	0.0000	{0.0000, 1.0000, 0.0000, 0.0000, 0.0000}
		2	1.0000	0.2000	{0.0000, 0.9200, 0.0800, 0.0000, 0.0000}
		3	1.0000	0.4000	{0.0000, 0.6800, 0.3200, 0.0000, 0.0000}
		4	1.0000	0.6000	{0.0000, 0.2800, 0.7200, 0.0000, 0.0000}
		5	1.0000	0.8000	{0.0000, 0.0000, 0.7200, 0.2800, 0.0000}
		6	1.0000	1.0000	{0.0000, 0.0000, 0.0000, 1.0000, 0.0000}

(2) Utility values used for the referential values of consequent attribute, e.g. five utility values are defined as follows:

$$\{u(D_n), n = 1, \ldots, 5\} = \{-0.5,\ 0,\ 0.5,\ 1,\ 1.5\}. \tag{7}$$

The belief rule can be written as below:

$$R_k : \textit{IF } U_1 \textit{ is } A_1^k,\ \textit{THEN } D \textit{ is } \{(D_n, \beta_{n,k}); n = 1, \ldots, 5\},\ \textit{with rule weight } \theta_k \textit{ and antecedent attribute weight } \delta, k = 1, 2, \ldots, L \tag{8}$$

To illustrate the generalization capability of the BRBS under different numbers of rules, three kinds of BRBSs with 1, 2, and 6 rules (i.e., $L = 1$, 2, and 6 respectively) can be constructed based on Section 2.2, named as BRBS-1, BRBS-2, and BRBS-6, respectively. Based upon the point of view that the training error of the BRBS on the training data is kept as small as possible, the trained parameters of BRBS-1, BRBS-2, and BRBS-6 are listed in Table 1, which are achieved through the following ways respectively:

(1) For BRBS-1, the inference outputs are fixed for all input data because only one belief rule can be activated. Hence, according to Appendix B, the ideal inference output of BRBS-1 is 0.3667 which can make sure the minimal training error of BRBS-1 on the training data. As a result, the distributed belief degrees for the consequent in the belief rule can be generated by using the utility-based equivalence transformation technique [48]. It means that 0.3667 is equivalent to {(D_2, 0.2667), (D_3, 0.7333)}. Other parameters, including antecedent attribute weight and rule weight, are all assumed to be 1 because these parameters have no effect on the inference output of BRBS-1.

(2) For BRBS-2, it has been investigated for illustrating the distributed approximation process of the BRBS [47]. Though parameter optimization, the mean square error (MSE) of the BRBS is reached to 9.7×10^{-5} and it can be reduced further if more referential values of the consequent attribute are used. For convenience, the trained BRBS found in [47] is regarded as BRBS-2.

(3) For BRBS-6, the ideal result of parameter optimization can be described that arbitrary one training input data is only to activate one belief rule and the inference output of the BRBS is equal to the training output data. For example, while the training input data is 0.2, the activated weights for all belief rules are 0 but the second belief rule being 1. Thus, the deviation between the inference output of BRBS-6 and the training output data is equal to 0. By using the utility-based equivalence transformation technique, the distributed belief degrees for the consequent in those six belief rules can be generated directly from those six training data.

To compare the generalization capability of three BRBSs shown in Table 1, a set of 100 data in the interval [0, 1] is generated uniformly. Table 2 and Fig. 4, in different ways, show the comparison results among three BRBSs in approximating the training and testing data sets. From Table 2, the mean absolute error (MAE) for the training data set is decreasing from 0.2889 to 0.0000 progressively while the number of rules is increasing, in which the MAE of the BRBS is reached to zero when there are six belief rules. However, the MAE in the testing data set shows the results opposite to the original expectations that BRBS-2 has the minimal MAE better than BRBS-1 and BRBS-6. From Fig. 4, it is clear that (1) too few rules have direct influences on the generalization capability of the BRBS in both the training and the testing data sets; (2) although BRBS-6 seems to be performing well in approximating the training data set, the priority of BRBS-2 is superior to that of BRBS-6 in the testing data set.

Table 2
Comparison of MAE for three BRBs in terms of accuracy.

Type	No. of rules	Training data set		Testing data set	
		MAE	Priority	MAE	Priority
BRBS-1	1	0.2889	3	0.2651	3
BRBS-2	2	0.0051	2	0.0087	1
BRBS-6	6	0.0000	1	0.0149	2

The analytical and numerical analyses of this section reveal the generalization capability of the BRBS under different numbers of rules, leading to a conclusion that a BRBS may have a perfect accuracy in the training data set, but it does not mean the BRBS is available to generate the ideal inference output for the testing data set. Thus, a new criterion should be proposed to evaluate the generalization capability of the BRBS while using the training data set to obtain the optimal parameters and structure of a BRB.

3.2. Generalization error of BRBS

In this section, it is further illustrated the generalization capability of the BRBS by using the theorem of Hoeffding inequality [36], which provides an upper bound on the probability that the sum of random variables deviates from its expected value. To date, the theorem of Hoeffding inequality has been introduced to investigate the generalization ability of traditional models, such as artificial neural networks [44], Bayesian model [45] and support vector machine [46].

Theorem 1. ***Hoeffding Inequality Theorem***[36]: *Let $X_1,\ldots,X_n$ be independent random variables bounded by the interval $a_i \le X_i \le b_i$ ($i = 1,\ldots,n$), the mean of these variables is denoted by*

$$\bar{X} = \frac{X_1 + \cdots + X_n}{n}. \tag{9}$$

Then for $t > 0$, the Hoeffding inequality can be written as follow:

$$P\left(\bar{X} - E\bar{X} \ge t\right) \le \exp\left(-\frac{2n^2t^2}{\sum_{i=1}^{n}(b_i - a_i)^2}\right), \tag{10}$$

where $E\bar{X}$ denotes the expected value of the mean $\bar{X}$.

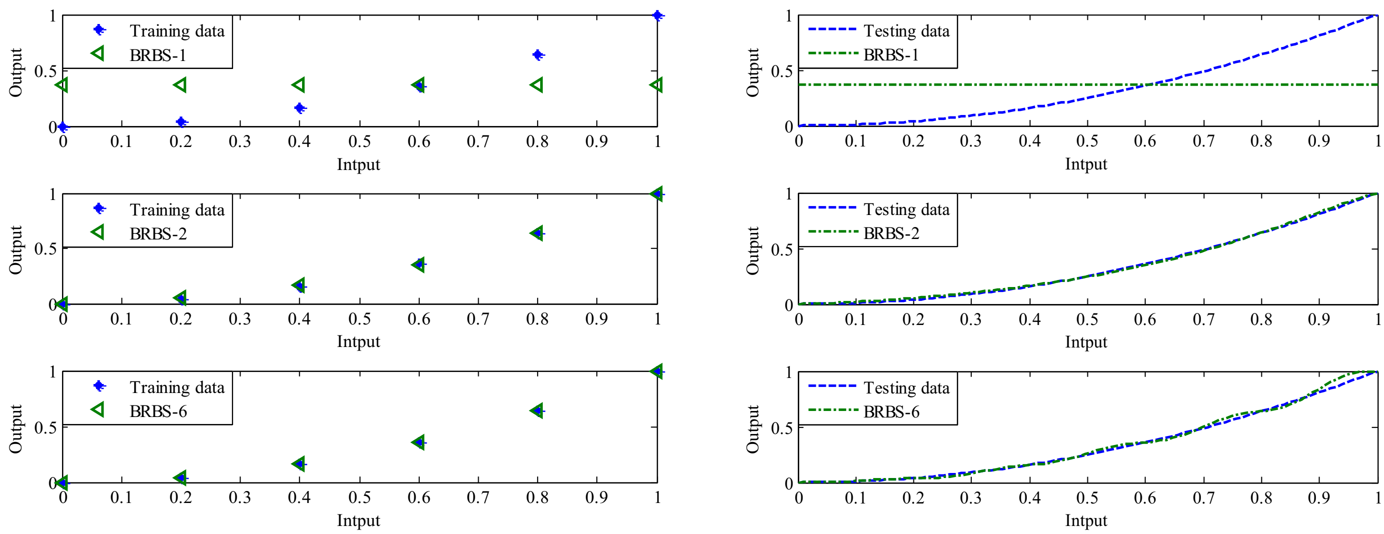

Fig. 4. Three approximation curves on the training and the testing data.

In the following, we will set up the link of the BRBS and the Hoeffding inequality theorem.

Let $\boldsymbol{x}=(x_1, x_2,..., x_M)$ be a M-ary variable, J_i be the number of the referential values for the ith antecedent attribute ($i=1, 2,..., M$) and d be the total number of the referential values for all antecedent attributes ($d=\sum_{i=1}^{M} J_i$). Based on these referential values and suppose only one consequent attribute D with N referential values D_n ($n=1,..., N$), different BRBs can be constructed which leads to different BRBSs. Assume $f_s(\boldsymbol{x})$ ($s=1,..., S$) is the numerical inference output of the sth BRBS and also represents the analytical model of the sth BRBS by using the unified inference procedure of the BRBS given in Eqs. (A1)–(A8) (detailed in Appendix A), here S is the total number of BRBSs constructed from those d referential values. Let $F(\boldsymbol{x})$ be the set of all analytical models of BRBSs and the size of $F(\boldsymbol{x})$ is $S=2^d$. For example, if one antecedent attribute with two referential values, the number of possible BRBS options would be 4.

In the following, to simplify expressions, some intermediate variables are defined as follows:

$$\chi_n^1(\boldsymbol{x})=\prod_{k=1}^{L}\left(w_k\beta_{n,k}+1-w_k\sum\nolimits_{i=1}^{N}\beta_{i,k}\right)-\prod_{k=1}^{L}\left(1-w_k\sum\nolimits_{i=1}^{N}\beta_{i,k}\right), \tag{11}$$

$$\chi^2(\boldsymbol{x})=\prod_{k=1}^{L}\left(1-w_k\sum\nolimits_{i=1}^{N}\beta_{i,k}\right)-\prod_{k=1}^{L}(1-w_k). \tag{12}$$

Based on Eqs. (11) and (12), the combined belief degree expressed in Eq. (A7) and the numerical inference output expressed in Eq. (A8) in Appendix A can be re-expressed.

Lemma 1. *Each BRBS analytical model $f(\boldsymbol{x})\in F(\boldsymbol{x})$ can be regarded as a bounded independent random variable giving the input variable $\boldsymbol{x}$. In other words, the probability distribution of the random variable $f(\boldsymbol{x})$ is a list of probabilities associated with each of its possible values, that is, the complete set of referential values D_i ($i=1,..., N$).*

Proof. Assume $\{D_1,..., D_N\}$ is the complete set of the referential values for the consequent D. By analyzing integration of the activated belief rule and Eq. (A7) in Appendix A, the probability distribution of $f(\boldsymbol{x})$ is given as below:

$$P(f(\boldsymbol{x})=D_n)=\beta_n=\frac{\chi_n^1(\boldsymbol{x})}{\sum_{i=1}^{N}\chi_i^1(\boldsymbol{x})-\chi^2(\boldsymbol{x})},\ n=1,\dots,N. \tag{13}$$

Hence, each BRBS with a given input variable can be treated as a random variable. In addition, for each $f(\boldsymbol{x})\in F(\boldsymbol{x})$, based on Eq. (A8) in Appendix A, $f(\boldsymbol{x})$ is bounded by the interval $[u(D_1), u(D_N)]$ [47], in which $u(D_1)$ is the lowest utility value of the referential value and $u(D_N)$ is the highest utility value of the referential value regarding the consequent attribute.

In addition, assume two independent input data $\boldsymbol{x_1}$ and $\boldsymbol{x_2}$ are provided for $f(\boldsymbol{x})$, by analyzing calculation of the individual matching degree and the activated weight in Appendix A, two independent input data can generate two independent activated weights for each belief rule. As a result, the combined belief degree on the nth referential value for the input data $\boldsymbol{x_1}$ and $\boldsymbol{x_2}$ can be expressed as follows:

$$\begin{aligned}P(f(\boldsymbol{x_1})=D_n,\, f(\boldsymbol{x_2})=D_n)\\ &=\frac{\chi_n^1(\boldsymbol{x_1})\chi_n^1(\boldsymbol{x_2})}{(\sum_{i=1}^{N}\chi_i^1(\boldsymbol{x_1})-\chi^2(\boldsymbol{x_1}))(\sum_{i=1}^{N}\chi_i^1(\boldsymbol{x_2})-\chi^2(\boldsymbol{x_2}))}\\ &=P(f(\boldsymbol{x_1})=D_n)P(f(\boldsymbol{x_2})=D_n).\end{aligned} \tag{14}$$

Hence, the inference output of the BRBS is independent while giving the independent input data.

By using the Hoeffding inequality theorem together with Lemmas 1 and boundedness property of BRBS [47], a quantitative criterion named as generalization error can be used to describe the generalization capability of BRBS and is concluded as follows:

Theorem 2. For each $f(\boldsymbol{x})\in F(\boldsymbol{x})$, while T independent input-output data pairs $(\boldsymbol{x_i}, y_i)$ ($i=1,...,T$) are provided, $f(\boldsymbol{x_1}),..., f(\boldsymbol{x_T})$ can be independent random variables bounded by the interval $u(D_1)\le f(\boldsymbol{x_i})\le u(D_N)$ ($i=1,...,T$) and there exists the following inequality:

$$ER(f)<R(f)+\sqrt{\frac{(d\ln 2-\ln r)(u(D_N)-u(D_1))^2}{2T}}, \tag{15}$$

where $ER(f)$ denotes the generalization error of $f(\boldsymbol{x})$, $R(f)$ denotes the mean error of $f(\boldsymbol{x})$, d denotes the total number of the referential values used for all antecedent attributes, r denotes the probability value bounded by the interval (0, 1].

Proof. See Appendix C.

According to Theorem 2, it can be concluded that the generalization capability of the BRBS is depended on the mean error and bias, in which the generalization error is equivalent to the testing error, the mean error is equivalent to the training error, and the

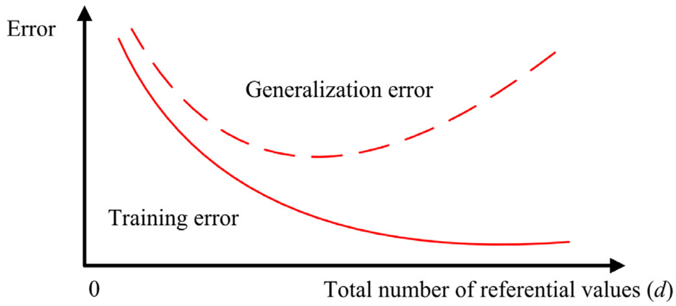

Fig. 5. Illustration of the generalization error and the training error changing with total number of referential values.

Table 3
Comparison of generalization error for three BRBs.

Type	BRBS-1	BRBS-2	BRBS-6
Generalization error	0.7885	0.5595	0.7338
Priority	3	1	2

bias is related to the total number of referential values and also the number of training data.

Remark 1. As illustrated in Section 2.2, the belief rule is constructed by linking the referential values in different antecedent attributes. In other words, the number of rules is related to the total number of referential values. Hence, based on Theorem 2, a BRBS with large number of referential values, namely rules, may achieve the low training error on the set of training data, but it might not achieve the low generalization error because the value of bias will be very large. The relationship between the generalization error and the training error can be illustrated in Fig. 5.

It is evident from Fig. 5 that the training error does not suit for determining the optimal structure of a BRB, but the generalization error does. For example, Table 3 shows the approximation generalization error of BRBS-1, BRBS-2, and BRBS-6 while assuming $r=0.1$. The priority of generalization capability of three BRBSs are BRBS-2 > BRBS-6 > BRBS-1 based on the generalization error of those BRBSs, whose priority order of three BRBSs is the same as that shown in Table 2. Therefore, the generalization error will be used in Section 4 as an effective measurement to help determine the good or acceptable structure of the BRB.

4. Joint optimization method on parameter and structure for BRBS

As a result from earlier discussions in Section 3, both the optimal parameters and the optimal structure of the BRB can contribute to construct an optimal BRBS, which can help to achieve better performance not only for the training data set, but also for the testing data set. Thus, in this section, two algorithms to optimize the parameters and structure of a BRB are proposed respectively, followed by a new generic algorithm to obtain the optimal parameters and structure of the BRB simultaneously.

4.1. Heuristic strategy for optimizing the structure of BRB

The optimal structure of a BRB, normally refer to the optimal number of parameters, plays an important role in generating an optimal BRBS and two issues need to be addressed.

Firstly, an evaluation criterion should be orientated into determining the optimal structure of the BRB. In the previous studies [34,35], the training error is always applied in structure optimization as a criterion to judge whether the structure of the BRB should be optimized or not. However, the training error may not be a good choice of criterion based on the investigation in Section 3. Hence, a definition of generalization error for the BRBS is provided as follows:

Definition 1. Suppose that a BRBS, whose number of antecedent attribute is M with J_i referential values for each antecedent attribute, namely, total number of the referential values used for all antecedent attributes is $d=\sum_{i=1}^{M} J_i$, and the inference output $f(\boldsymbol{x})$ is bounded by the interval $[u(D_1), u(D_N)]$, has been trained by using T independent input-output data pairs $(\boldsymbol{x_i}, y_i)$ $(i=1,\ldots,T)$, the generalization error of the BRBS is defined as follows:

$$ge = \frac{\sum_{i=1}^{T} |f(\boldsymbol{x_i}) - y_i|}{T} + \sqrt{\frac{(d\ln 2 - \ln r)(u(D_N) - u(D_1))^2}{2T}}, \quad (16)$$

where $f(\boldsymbol{x_i})$ is the inference output of the BRBS generated by the ith input data, y_i is the actual output of the ith input data; r is the probability value bounded by the interval (0, 1].

Remark 2. The probability value r is crucial to calculate the generalization error of a BRBS because it represents the probability to obtain Eq. (15), which is the basis of Definition 1. Based on Eq. (C9) shown in Appendix C, if r is small, it is more likely to use Definition 1 for describing the generalization capability of the BRBS. However, it would weaken the function of training error for the generalization error of the BRBS. Hence, the value of r should be set according to the actual situation.

Secondly, an efficient strategy to optimize the structure of BRB is expected. Normally, the number of antecedent attributes and the referential values are fixed based on the application context, the number of rules is closely linked to the number of the referential values for each antecedent attribute. So, the latter is used as a concise scheme to express the structure of a BRB in the following algorithms. However, it is not easy to obtain the optimal structure of the BRB because there are too many possible BRBs. For example, assume there are two antecedent attributes and each attribute has three, four, and five referential values, respectively. As a result, owing to any referential value likely being used for constructing a BRB, a total of $2^6=64$, $2^8=256$ and $2^{10}=1024$ possible BRBs have to be considered in the structure optimization. This is a combinatorial explosion problem: the kind of BRBs would grow exponentially along with the increase of the number of referential values. Therefore, a heuristic strategy needs to be developed and it is provided as follows:

Definition 2. Suppose that there are two BRBSs with the optimized parameters and the different structures of the BRB, the heuristic strategy of structure optimization is then defined as that the number of referential values for each antecedent attribute in the BRBS, which has a smaller generalization error, is regarded as the tendency to add or prune that in another BRBS.

Based on Definitions 1 and 2, the structure optimization shown in Fig. 6 is described that two BRBs expressed by the optimized parameters under the different structures are provided to generate the generalization error, where the structure of the BRB is expressed as a concise scheme based on the referential values of the BRB. After adding/pruning referential values for each antecedent attribute according to the heuristic strategy, the structure of a new BRB can be generated. As a result of the previous discussion, the procedure of structure optimization can be summarized in the following algorithm:

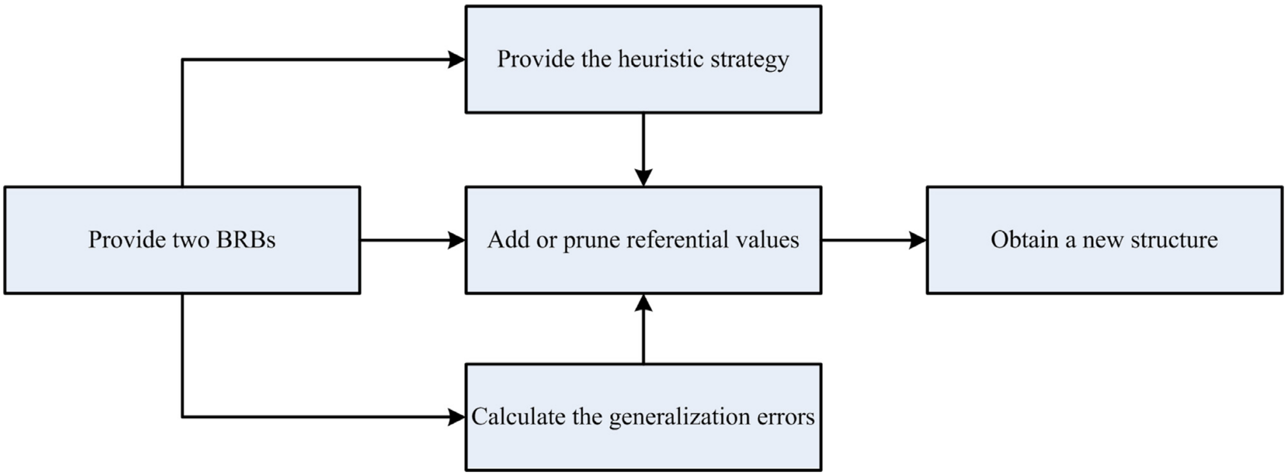

Fig. 6. Flowchart of structure optimization.

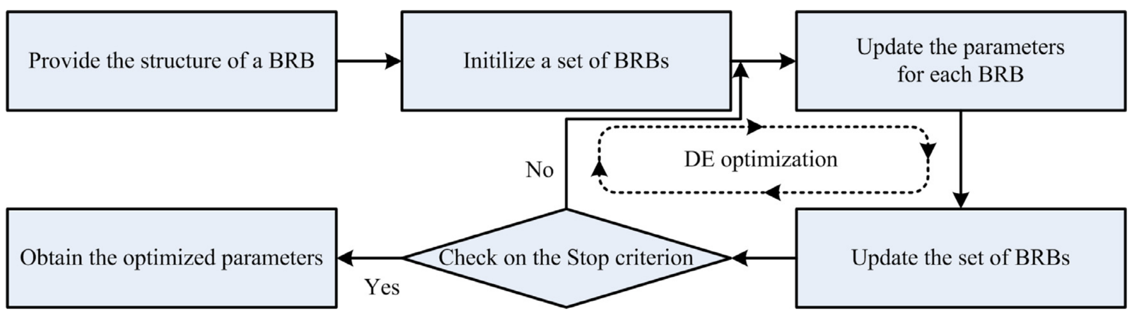

Fig. 7. Flowchart of parameter optimization.

SOHS algorithm: an algorithm to achieve **S**tructure **O**ptimization based on the **H**euristic **S**trategy for a BRBS

Input: Ω denotes the set of BRBs expressed by the optimized parameters under the different structures, where Ψ^k and Φ^k denote the optimized parameters and the structure of the kth BRB, respectively; φ denotes the probability value.

Output: The structure of a new BRB (Φ^{K+1}).

Begin
$\Phi^{K+1} = \{\}$.
Randomly select two individuals from Ω, denoted as $<\Psi^k, \Phi^k>$ and $<\Psi^l, \Phi^l>$, respectively.
Calculate the generalization errors based on Ψ^k and Ψ^l, denoted as ge^k and ge^l, respectively.
If $ge^k > ge^l$ then
  Exchange all elements of $<\Psi^k, \Phi^k>$ and those of $\langle \Psi^l, \Phi^l \rangle$.
Else if
For each antecedent attribute U_i in the attribute set U do
  Count the number of referential values for U_i based on Φ^k and Φ^l, denoted as J_i^k and J_i^l, respectively.
  If $J_i^k \geq J_i^l$ then
    $\Phi^{K+1} = \Phi^{K+1} \cup \Phi^l$.
    For $j = 1$ to J_i^l do
      If $Random(0, 1) < \varphi$ then
        Generate a new referential value $A_{i,j}^{K+1}$.
        $\Phi^{K+1} = \Phi^{K+1} \cup \{A_{i,j}^{K+1}\}$.
      End if
    End for
  Else then
    For each referential value $A_{i,j}^l$ in Φ^l
    If $Random(0, 1) >= \varphi$ then
      $\Phi^{K+1} = \Phi^{K+1} \cup \{A_{i,j}^l\}$.
    End if
    End for
  End if
End for
End

For the above mentioned SOHS algorithm, the following remarks are given:

Remark 3. For the structure of a new BRB, it should have twice or half numbers of referential values for each antecedent attribute comparing to two given BRBs based on Definition 2. Meanwhile, it is necessary to consider the diversity of the BRB. Thus, the addition or pruning of the referential value is implemented by the random way.

Remark 4. The pre-defined probability φ is crucial for the SOHS algorithm. Obviously, if φ is too small, it is possible for a small difference to generate the structure of a new BRB, which contributes to search for the optimal structure of the BRB in a more precise way, but it is also more complex. In general, the probability φ is set to around 2^{-M} because the total probabilities of generating a new BRB is not less than 100% under the undesirable condition that there are only two referential values for each antecedent attribute, where M denotes the number of antecedent attributes in a BRB.

4.2. Differential evolution algorithm for optimizing the parameters of BRB

Once the structure of a BRB is generated by using the SOHS algorithm, all parameters of the BRB needs to be optimized by using the training data set. In this section, an optimization technique based on the differential evolution (DE) algorithm [37] is proposed to achieve the parameter optimization. The illustration of the optimization technique is shown in Fig. 7.

As shown in Fig. 7, the structure of a BRB is provided to construct a set of BRBs with different initial parameters. Then DE optimization and the training data set are used to iteratively optimize the parameters of each BRB. While the stop criterion is met, the optimized parameters of the BRB are generated under the given structure of the BRB. The procedure of parameter optimization can be summarized in the following algorithm:

PODE algorithm: an algorithm to achieve **P**arameter **O**ptimization based on **DE** algorithm for a BRBS

Input: Φ denotes the structure of a BRB; ε and S_1 denote the stop criterions of DE algorithm; F denotes the mutation operator of the DE algorithm; CR denotes the crossover operator of the DE algorithm; Ψ^k denotes the parameters of the kth BRB, and p_i^k denotes the ith parameter in Ψ^k, including rule weights, antecedent attribute weights, belief degrees, and utility values of referential values used for antecedent attributes.

Output: The optimized parameters of the BRB (Ψ^{K+1}).

Begin
Randomly generate the initial parameters of $K+1$BRBs based on the structure Φ, denoted as $\Psi^1, \ldots, \Psi^{K+1}$, respectively.
While the variation value of the generalization error for Ψ^{K+1} is smaller than ε in the latest S_1 iterations do
  For each Ψ^k in $\{\Psi^1, \ldots, \Psi^K\}$ do
    Randomly select three different individuals from $\{\Psi^1, \ldots, \Psi^K\}$, denoted as Ψ^{k_1}, Ψ^{k_2}, and Ψ^{k_3} respectively.
    Randomly generate the initial parameters of a new BRB based on the structure Φ, denoted as Ψ^{k_0}.
    For each parameter $p_i^{k_0}$ in Ψ^{k_0} do
      $p_i^{k_0} = p_i^{k_1} + F * (p_i^{k_2} - p_i^{k_3})$.
      If $Random(0, 1) < CR$ then
        $p_i^{k_0} = p_i^k$.
    End if
    If $p_i^{k_0}$ does not meet the constraints (such as Eqs. (2a)–(2f) in [30] and Eqs. (4a)–(4c) in [40]) then
        Randomly generate a new parameter for $p_i^{k_0}$ based on the given constraint.
      End if
    End for
    Calculate the generalization errors based on Ψ^{k_0}, Ψ^k and Ψ^{K+1}, denoted as ge^{k_0}, ge^k and ge^{K+1}, respectively.
    If $ge^{k_0} \le ge^k$ then
      Replace all parameters of Ψ^k using those of Ψ^{k_0}.
    End if
    If $ge^k \le ge^{K+1}$ then
      Replace all parameters of Ψ^{k+1} using those of Ψ^k.
    End if
  End for
 End while
End

For the above mentioned PODE algorithm, the following remarks are given:

Remark 5. In previous optimization techniques, such as [32] and [43], the stop criterion based on the number of iterations is unable to avoid the unwished situation that there are too many/few iterations in the parameter optimization. Thus, the stop criterion of the PODE algorithm is based on the variation value of the generalization error within the latest iterations, which can ensure the reasonable number of iterations.

Remark 6. Different parameter optimization models may have different constraints for the same parameter, such as Eqs. (2a)–(2f) of local parameter optimization model [30] and Eqs. (4a)–(4c) of model for the BRB under disjunctive assumption [40]. However, the PODE algorithm is a generic optimization technique that can be used to achieve the parameter optimization under different models shown in Sections 5.1.2, 5.1.3, and 5.2.2, respectively.

Remark 7. In the PODE algorithm, the thresholds ε and S_1 are set according to the desired accuracy. The thresholds F and CR are set to around 0.5 and 0.9 based on the empirical experience from different case studies [32,43].

4.3. Joint optimization method for optimizing the parameters and structure of BRB

To optimize the parameters and structure of the BRB simultaneously, a joint optimization method is proposed to achieve parameter and structure optimization by integrating the SOHS algorithm and the PODE algorithm. As shown in Fig. 8, on the one hand, the structure of a new BRB can be generated by using the SOHS algorithm and the set of BRBs expressed by the optimized parameters under the different structures. On the other hand, the optimized parameters of the BRB can be generated by using the PODE algorithm under the given structure of the BRB and then are used to update the set of BRBs expressed by the optimized parameters under the different structures. After iteratively achieving parameter and structure optimization, it is finally to obtain the optimal parameters and structure of the BRB.

However, not all BRBs, which have been optimized by using the SOHS and PODE algorithms, are available to update the set of BRBs expressed by the optimized parameters under the different structures because some BRBs with a bigger generalization error and the more number of referential values have been impossible to be the optimal parameters and structure of the BRB. As such, the Pareto frontier [49] is introduced to screening those unnecessary BRBs. The screening strategy is defined as follows:

Definition 3. Let Ω be the set of BRBs expressed by the optimized parameters under the different structures, where Ψ^k and Φ^k denote the optimized parameters and the structure of the k ($k=1,\ldots,|\Omega|$) th BRB. By using Ψ^k and Φ^k, it is easy to obtain the number of referential values for each antecedent attribute $J_i^k(i=1,\ldots,M)$ and the generalization error ge^k. The screening strategy is then defined as that $<\Psi^t, \Phi^t>$ is unnecessary to update the set Ω if and only if arbitrary an individual $<\Psi^k, \Phi^k> \in \Omega$ meets the following conditions:

$$\begin{cases} J_i^k \le J_i^t;\ i = 1, \ldots, M \\ ge^k \le ge^t \end{cases} \tag{17}$$

Based on Definition 3 together with the SOHS algorithm and the PODE algorithm proposed above, a joint optimization method on parameter and structure is summarized in the following algorithm:

JOPS algorithm: an algorithm to achieve the **J**oint **O**ptimization on **P**arameter and **S**tructure for a BRBS.

Input: Φ^0 denotes an initial structure of the BRB and S_2 denote the stop criterion of the iteration process.

Output: The optimal parameters and structure of the BRB ($<\Psi^{K+1}, \Phi^{K+1}>$).

Begin
Obtain the optimized parameters of the BRB using the PODE algorithm and Φ^0, denoted as Ψ^0.
$\Omega = \{<\Psi^0, \Phi^0>\}$.
While the set Ω does not change in the latest S_2 iterations do
  Obtain the structure of a new BRB using the SOHS algorithm and Ω, denoted as Φ^t.
  Obtain the optimized parameters of the BRB using the PODE algorithm and Φ^t, denoted as Ψ^t.
  Update the set Ω using $<\Psi^t, \Phi^t>$ based on Definition 3.
End while
Select arbitrary an individual from Ω, denoted as $\langle\Psi^{k+1}, \Phi^{k+1}\rangle$.
For each $<\Psi^k, \Phi^k>$ in the set Ω do
  Generate the generalization errors ge^k and ge^{K+1} using Ψ^k and Ψ^{K+1}.
  If $ge^k < ge^{K+1}$ then
    Replace all elements of $<\Psi^{K+1}, \Phi^{K+1}>$ by those of $\langle\Psi^k, \Phi^k\rangle$.
   End if
  End for
End

For the above mentioned JOPS algorithm, the following remarks are given:

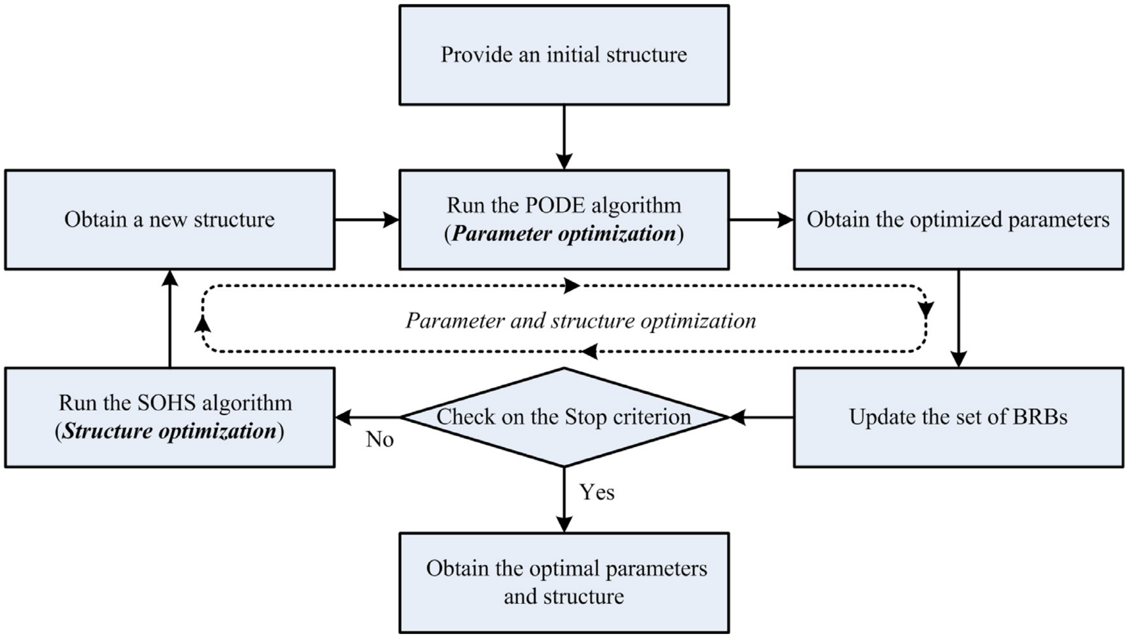

Fig. 8. Flowchart of the parameter and structure optimization.

Remark 8. For a complex system, prior or expert knowledge may not be perfect to construct a complete structure of the BRB from the beginning. However, a simple initial structure of the BRB, which just has the least preferred referential values and most preferred referential values for each antecedent attribute, can be easily constructed and is enough to be the input of the JOPS algorithm.

Remark 9. In the present work, the set of referential values for each antecedent attribute is used as a concise scheme to express the structure of the BRB. For a given set of referential values, the structure of the BRB under conjunctive assumption can be constructed based on [15] and that under disjunctive assumption can be constructed according to [41].

Remark 10. In the JOPS algorithm, the threshold S_2 is set according to the desired accuracy. If the S_2 is too large, the performance of the EBRBS will be better, but the running time of the JOPS algorithm will be worse.

5. Numerical studies

In this section, two practical problems, namely oil pipeline leak detection and bridge risk assessment, are applied to validate the proposed joint optimization method and reveal its practicability. Each numerical study is firstly to show the background of practical problem and then to perform the proposed method on the BRB under conjunctive and disjunctive assumptions. Finally, the existing studies are introduced to compare with the proposed method.

5.1. Joint optimization for oil pipeline leak detection

5.1.1. Background formulation

A series of leak data were collected from a 100 km oil pipeline installed in Great Britain. The leak data includes flow difference (FD), pressure difference (PD) and leak size (LS), respectively. FD and PD are the two important factors in detecting whether there is a leak in the pipeline and are related to LS. Therefore, FD and PD are selected as the antecedent attributes of the BRB, and LS is considered as the consequent attribute for the BRB.

During a leak trial period, 2008 samples of 25% leak data were collected at a rate of 10 per samples, as shown in Fig. 9. To validate the proposed method, we select four periods, including 7:00–7:49 a.m., 8:50–9:30 a.m., 9:40–10:10 a.m., and 10:30–11:20 a.m., to collect 900 training data. Furthermore, according to the previous work [17], only two referential values, i.e. negative large (NL) and positive large (PL) are provided for FD and PD and their utility values are shown respectively as follows:

$$(A_{FD}, u(A_{FD})) \in \{(NL, -10), (PL, 3)\}. \tag{18}$$

$$(A_{PD}, u(A_{PD})) \in \{(NL, -0.02), (PL, 0.02)\}. \tag{19}$$

Five referential values used for the consequent attribute, including zero (Z), very small (VS), medium (M), high (H), and very high (VH), are provided for LS and their utility values are given as follows:

$$(D_{LS}, u(D_{LS})) \in \{(Z, 0), (VS, 2), (M, 4), (H, 6), (VH, 8)\}. \tag{20}$$

To validate the generality of the proposed method for different parameter optimization models, the BRB under disjunctive and conjunctive assumptions are investigated in Sections 5.1.2 and 5.1.3, respectively. In addition, it is assumed that $r=0.1$ for calculating the generalization error, $\varphi = 0.25$ for running the SOHS algorithm, $\varepsilon=0.01$ and $S_1=200$ for running the PODE algorithm, and $S_2=3$ for running the JOPS algorithm.

5.1.2. Generality analysis on BRB under conjunctive assumption

For the BRB under conjunctive assumption, Fig. 10 shows the total number of referential values and the generalization error changes with the progress of iterations in the proposed method, in which the square denotes the value at the current iteration and the triangle denotes the best value collected from the starting to the current iteration. For example, from the starting to the 2nd iteration, the current iteration can obtain the result better than the previous one. Hence, there is an overlap between the square and triangle in Fig. 9. From the 3rd to the 5th iterations, the current iteration cannot obtain the result better than the 2nd iteration. Therefore, there is a line for the best result in Fig. 10. When the algorithm is stopped, the generalization error is reached to a fixed value 0.6498 and 5 for the number of referential values used for antecedent attributes. In the following, the intermediate results of

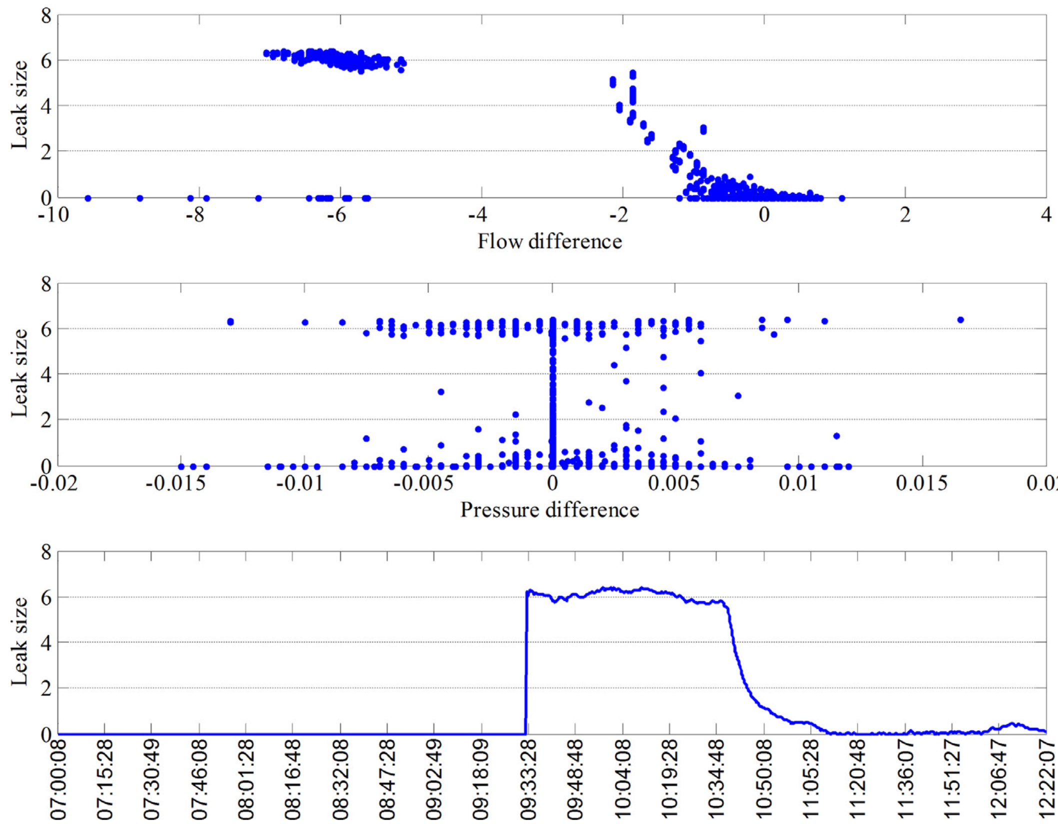

Fig. 9. 2008 samples of oil pipeline leak data.

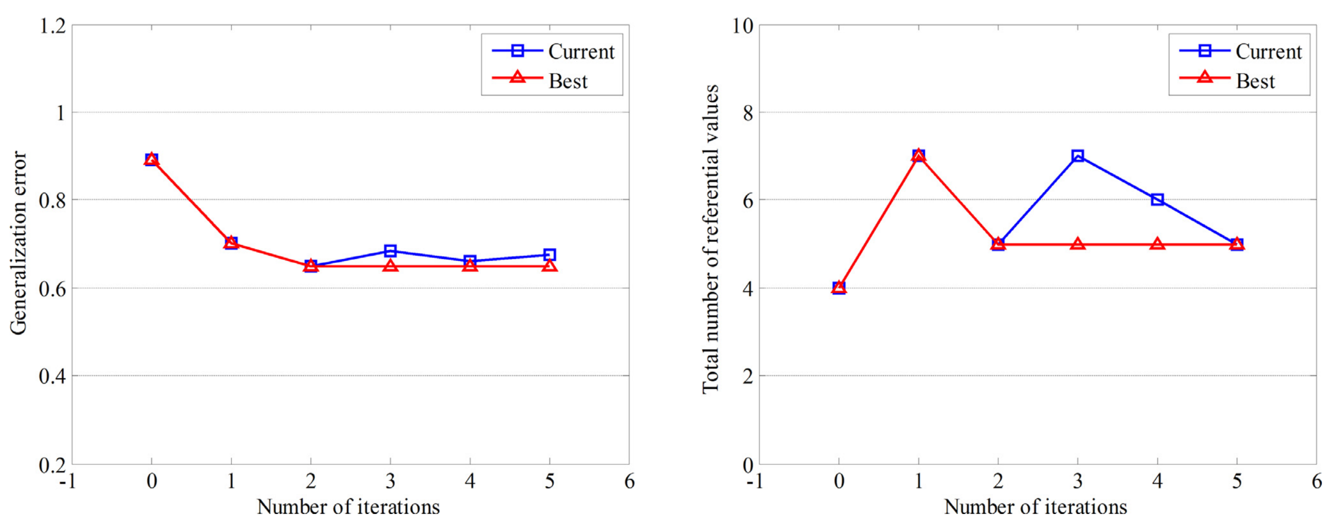

Fig. 10. Changes of the generalization error and the number of referential values with the five iterations.

the proposed method are introduced to show a more detailed procedure of optimizing the parameters and structure of the BRB.

At the beginning of the JOPS algorithm, the initial structure of the BRB, which has four belief rules constructed by using the given referential values shown in Eqs. (18) and (19), is used for the PODE algorithm. Table 4 shows the optimized parameters of the BRB and its generalization error is 0.8902, where the parameter optimization model is based on Eqs. (2a)–(2e) and (3). For the convenience of discussion, the generalization error is used to express the optimized parameters of the BRB and the numbers of referential values for each antecedent attribute are applied to denote the structure of the BRB. Thus, the trained BRB shown in Table 4 can be used to initialize the set Ω and it is shown as {⟨0.8902, (2, 2)⟩} finally.

Based on the above initialized set Ω, the structure of a new BRB is then generated using the SOHS algorithm. Fig. 11 illustrates the numbers of referential values regarding FD and PD for each iteration. Once it is done, the PODE algorithm is then used to optimize the parameters of the BRB constructed by the structure above. Finally, the optimized parameters of the BRB under the given structure should be used to update the set Ω. As shown in Fig. 10, there are five iterations to optimize both parameters and structure of the BRB, and the generalization error decreases in the first and second iterations. Taking the first iteration for example, the ⟨0.8902, (2, 2)⟩ in the set Ω generated based on the initial structure and the PODE algorithm can be only used to generate the structure of a new BRB, and two to four referential values may be generated with the prob-

Table 4
Trained BRB ⟨0.8902, (2, 2)⟩ with four rules.

Rule no.	Rule weight	Flow difference $\delta_1 = 1.0000$	Connective and	Pressure difference $\delta_2 = 0.2298$	Leak size
1	0.6362	−10.0000	and	−0.0200	{0.0001, 0.0001, 0.0001, 0.0003, 0.9993}
2	0.0000	−10.0000	and	0.0200	{0.1096, 0.2905, 0.2185, 0.2379, 0.1435}
3	0.8600	3.0000	and	−0.0200	{1.0000, 0.0000, 0.0000, 0.0000, 0.0000}
4	0.0000	3.0000	and	0.0200	{0.1807, 0.4093, 0.1114, 0.1629, 0.1356}

Table 5
Trained BRB ⟨0.6498, (3, 2)⟩ with six rules.

Rule no.	Rule weight	Flow difference $\delta_1 = 0.9620$	Connective and	Pressure difference $\delta_2 = 0.2764$	Leak size
1	0.1001	−10.0000	and	−0.0200	{0.0401, 0.0639, 0.0467, 0.5671, 0.2822}
2	0.0934	−10.0000	and	0.0200	{0.0488, 0.1086, 0.1406, 0.0829, 0.6191}
3	0.0177	−1.0955	and	−0.0200	{0.3381, 0.6194, 0.0255, 0.0162, 0.0008}
4	0.0003	−1.0955	and	0.0200	{0.1545, 0.2931, 0.3997, 0.0939, 0.0587}
5	0.9723	3.0000	and	−0.0200	{0.9998, 0.0000, 0.0000, 0.0000, 0.0001}
6	0.0005	3.0000	and	0.0200	{0.0946, 0.2123, 0.1309, 0.4039, 0.1582}

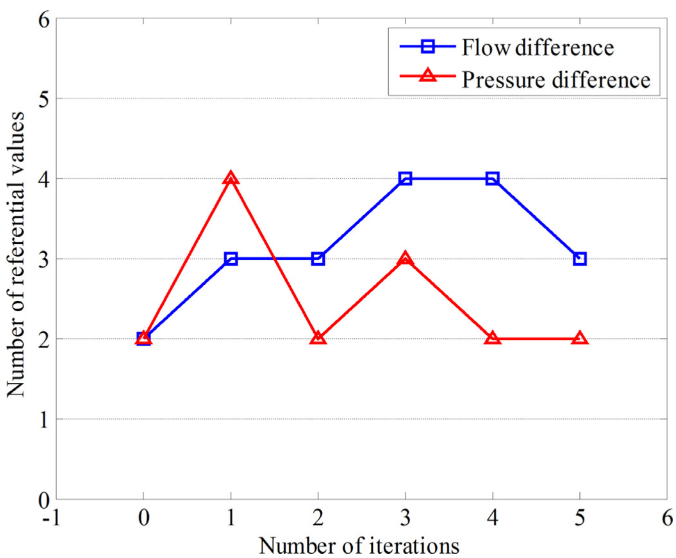

Fig. 11. Changes of referential values for FD and PD in five iterations.

ability 0.25 for each antecedent attribute. Finally, the new number of referential values regarding FD and PD are 3 and 4 after running the SOHS algorithm, and totally 12 belief rules can be constructed as the structure of a new BRB. The trained BRB ⟨0.7023, (3, 4)⟩ can be then generated using the PODE algorithm. Owing to Definition 3 that there exists 0.7023 < 0. 8902, the trained BRB should be put into the set Ω and the updated set is shown as {⟨0.8902, (2, 2)⟩, ⟨0.7023, (3, 4)⟩}.

At the ending of the JOPS algorithm, there are two trained BRBs in the set Ω after 5 iterations of parameter and structure optimization, namely, {⟨0.8902, (2, 2)⟩, ⟨0.6498, (3, 2)⟩}, whose optimized parameters are shown in Tables 4 and 5, respectively. As a result, ⟨0.6498, (3, 2)⟩ is selected as the optimal parameters and structure of the BRB because it has the minimal generalization error.

All of the 2008 data shown in Fig. 9 are used to validate the newly constructed BRB. As shown in Fig. 12, the estimated outcomes almost match the observed outcomes, in particular for the leak happening at around 9:33 a.m. to 10:47 a.m. However, it is also noticed that there is a big leak estimation error at the time around 9:32 a.m. to 9:33 a.m., which may be caused by the nosiy data.

5.1.3. Generality analysis on BRB under disjunctive assumption

For the BRB under disjunctive assumption, Fig. 13 shows the total number of referential values and the generalization error change with the iterative progress in the proposed method. Those values are reached to a fixed final value, namely 0.6420 and 5, respectively, while the algorithm is stopped. The following intermediate results are used to show the procedure. It is worth noting that each belief rule is not necessary to include all antecedent attributes because the BRB under disjunctive assumption is only required to cover all referential values once.

At the beginning of the JOPS algorithm, the referential values shown in Eqs. (18) and (19) are used to construct the initial structure of BRB and there are two belief rules in the BRB. After optimizing parameters using the PODE algorithm, the trained BRB is shown in Table 6 and its generalization error is 1.7496, in which Eqs. (2a)–(2e) and (4a)–(4c) are the parameter optimization model for the BRB under disjunctive assumption. Finally, the set Ω can be initialized to {⟨1.7496, (2, 2)⟩}.

Afterwards, the structure and also the optimized parameters of the BRB can be generated using the SOHS and PODE algorithms respectively through all iterations. Fig. 14 shows the numbers of referential values regarding FD and PD for each iteration. Taking the first iteration for example, based on the ⟨1.7496, (2, 2)⟩ in the set Ω, 4 and 2 referential values regarding FD and PD can be generated by using the SOHS algorithm with the probability 0.25, and the structure of a new BRB with 4 belief rules is then constructed. After running the PODE algorithm, the new BRB can obtain the optimized parameters and its generalization error is 0.6678. Owing to Definition 3 that there exists 0.6678 < 1.7496, the trained BRB should be put into the set Ω and the updated set is shown as {⟨1.7496, (2, 2)⟩, ⟨0.6678, (4, 2)⟩}.

At the ending of the JOPS algorithm, there are three trained BRBs in the set Ω after 6 iterations of parameter and structure optimization, namely, {⟨1.7496, (2, 2)⟩, ⟨1.0472, (2, 3)⟩, ⟨0.6420, (3, 2)⟩}, whose optimized parameters are shown in Tables 6–8, respectively. As a result, ⟨0.6420, (3, 2)⟩ is selected as the optimal parameters and structure of the BRB (Table 8).

All of the 2008 data shown in Fig. 9 are used to validate the BRB obtained by the proposed method. Fig. 15 displays the observed leak size and the estimated leak size on the time scale. It shows that the BRBS can clearly detect the leak which happened at around 9:36 a.m. and ended at around 10:47 a.m. and almost all absolute error is small but that around 9:32 a.m. to 9:33 a.m. Additionally, comparing with the estimated leak size shown in Fig. 12, there is little difference between the BRB under conjunctive and

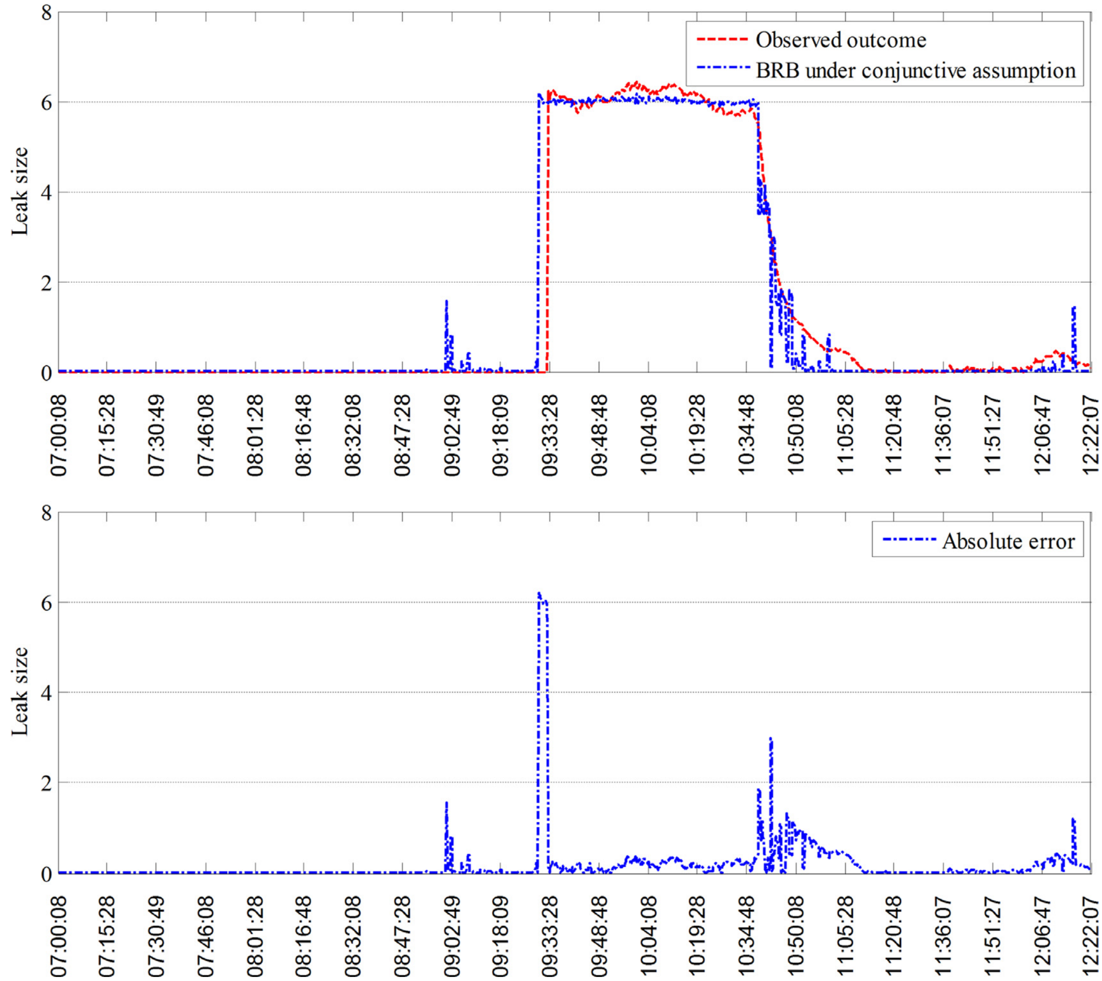

Fig. 12. Estimated outcomes and absolute error for the BRB under conjunctive assumption.

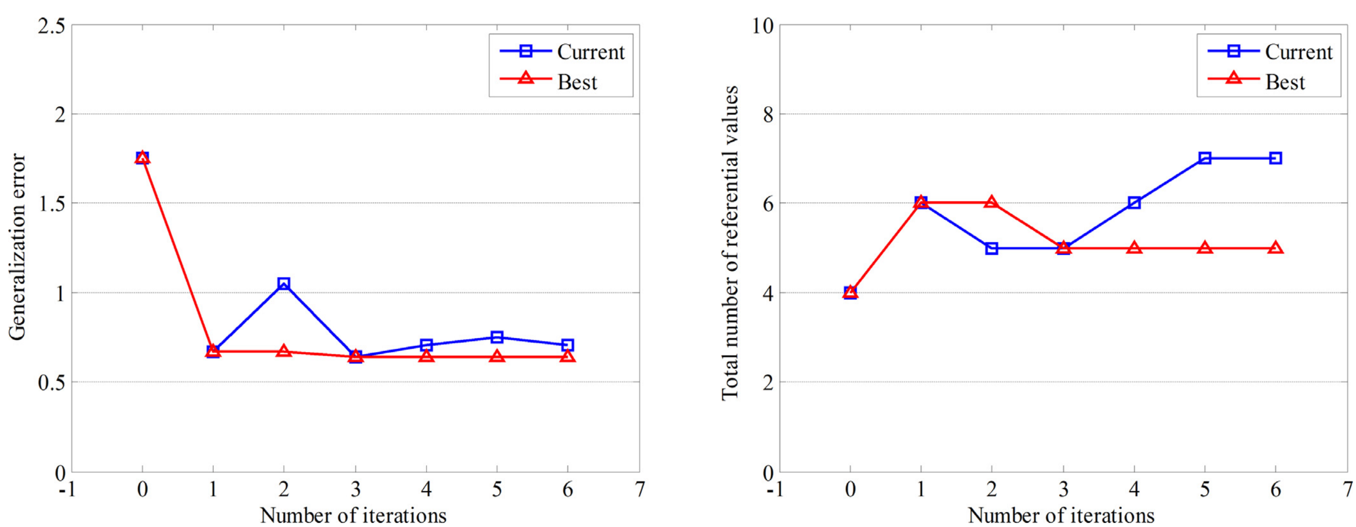

Fig. 13. Changes of the generalization error and the number of referential values with the six iterations.

Table 6
Trained BRB ⟨1.7496, (2, 2)⟩ with two rules.

Rule no.	Rule weight	Flow difference $\delta_1 = 1.0000$	Connective or	Pressure difference $\delta_2 = 0.9998$	Leak size
1	0.2623	−10.0000	or	−0.0200	{0.0000, 0.0000, 0.0001, 0.0001, 0.9998}
2	0.3717	3.0000	or	0.0200	{0.9999, 0.0000, 0.0001, 0.0000, 0.0000}

Table 7
Trained BRB ⟨1.0472, (2, 3)⟩ with three rules.

Rule no.	Rule weight	Flow difference $\delta_1 = 0.9999$	Connective or	Pressure difference $\delta_2 = 0.9996$	Leak size
1	0.6790	−10.0000	or	−0.0200	{0.0001, 0.0006, 0.0010, 0.0002, 0.9982}
2	0.9400	3.0000	or	0.0200	{0.9999, 0.0001, 0.0000, 0.0000, 0.0000}
3	0.0001	–	–	0.0000	{0.2066, 0.2171, 0.1538, 0.0617, 0.3607}

Table 8
Trained BRB ⟨0.6420, (3, 2)⟩ with three rules.

Rule no.	Rule weight	Flow difference $\delta_1 = 0.6937$	Connective or	Pressure difference $\delta_2 = 0.0283$	Leak size
1	0.0004	−0.7501	or	−0.0200	{0.3336, 0.1941, 0.0166, 0.2090, 0.2468}
2	0.1382	3.0000	or	0.0200	{1.0000, 0.0000, 0.0000, 0.0000, 0.0000}
3	0.6171	−10.0000	–	–	{0.0197, 0.0159, 0.2236, 0.0924, 0.6483}

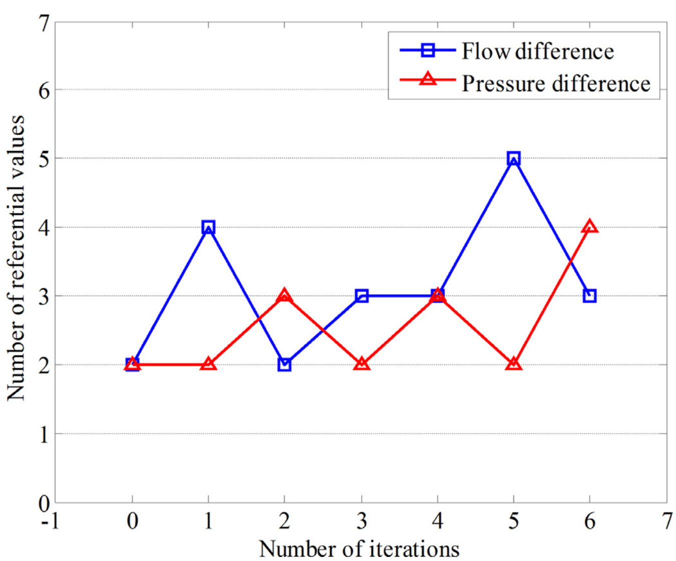

Fig. 14. Changes of referential values for FD and PD in six iterations.

disjunctive assumptions, but both of them produce small absolute error in almost 2008 data. In other words, it is believed that the proposed method can optimize the parameter and structure of the BRB under conjunctive and disjunctive assumptions.

5.1.4. Comparative analysis with some existing works

To further verify the validity of the proposed method, we have conducted a series of 20 runs of the JOPS algorithm for the BRB under conjunctive and disjunctive assumptions. Fig. 16 shows the average measured values, such as MAE, MSE, the generalization error (GE), the number of rules (NOR), and the numbers of referential values for FD (NORV-FD) and PD (NORV-PD).

By comparing the results of the BRB under conjunctive and disjunctive assumptions, it is clear that the latter is superior to the former in the terms of accuracies, including MAE, MSE, and GE. However, for the number of referential values, NORV-PD has fewer numbers in the BRB under conjunctive assumption and NORV-FD has fewer numbers in the BRB under disjunctive assumption. In addition, the BRB under disjunctive assumption is required to be a concise representation scheme for constructing belief rules so that the number of rules in the BRB under disjunctive assumption is significantly less than that in the BRB under conjunctive assumption.

Table 9 illustrates the comparison results concerning some existing works and the proposed method, where the GE and MAE of all studies except the sequential learning are obtained from 2008 data. For the BRB under the conjunctive assumption, although the achievement based on the sequential learning can achieve the minimal number of parameters and rules, the BRB is incomplete actually which means that it fails to activate belief rules for some special input data. Another achievement based on the dynamic rule adjustment, which is to find the best decision structure of a BRB, has the same number of parameters and rules to this study, it proves that the proposed method is also able to find the best decision structure for the BRB under conjunctive assumption. For the BRB under disjunctive assumption, the result obtained by the proposed method is superior to all previous studies in the BRB under conjunctive and disjunctive assumptions.

In summary, the comparison results have shown that the proposed joint optimization method can optimize the BRB in both conjunctive and disjunctive assumptions and provide superior inference performance, with the lower GE, MAE, less number of parameters and rules than other studies in oil pipeline leak detection.

5.2. Joint optimization for bridge risk assessment

5.2.1. Background formulation

Bridge risk assessment [38,39] is a well-known benchmark which has been used to validate many kinds of conventional methods such as the artificial neural network (ANN), evidential reasoning with learning (ERL), and multiple regression analysis (MRA). Moreover, comparing to oil pipeline leak detection, bridge risk assessment is a more complex decision problems because of more number of antecedent attributes shown in Fig. 17. Hence, it is believed that bridge risk assessment is sufficient to further confirm the performance of the proposed joint optimization method.

As shown in Fig. 17, the bridge risk is determined by four antecedent attributes, namely Safety, Functionality, Sustainability, and Environment. According to the data preprocessing strategy from [38], 66 and 506 bridge structures maintenance projects are extracted from the original 23, 387 projects to train and test the BRBS, respectively.

To construct the BRBS for bridge risk assessment, each antecedent attribute is provided two referential values with the utility values −1 and 4, respectively. Five ratings are used to define risk scores, namely *Zero* (*Z*), *Small* (*S*), *Medium* (*M*), *High* (*H*), and *Very High* (*VH*), and their utility values are expressed as follows:

$$D = \{Z, S, M, H, VH\} = \{0,\ 25,\ 50,\ 75,\ 100\} \tag{21}$$

According to the literatures [38,55], the root mean square error (RMSE), mean absolute percentage error (MAPE), and correlation coefficient (R) are used to evaluate the performance in assessing bridge risks. In addition, it is assumed that $r = 0.1$ for calculating the generalization error, $\varphi = 0.25$ for running the SOHS algorithm,

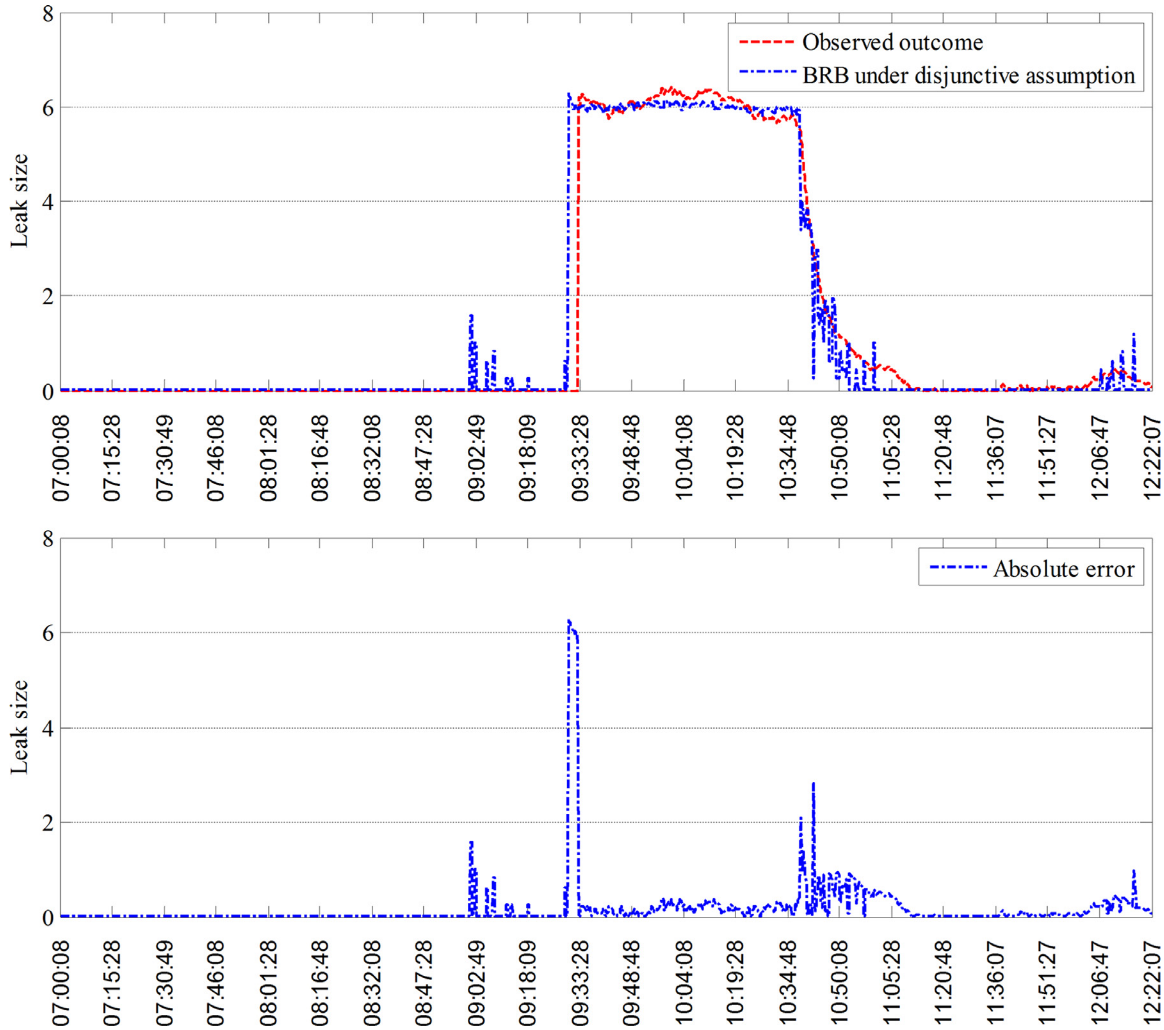

Fig. 15. Estimated outcomes and absolute error for the BRB under disjunctive assumption.

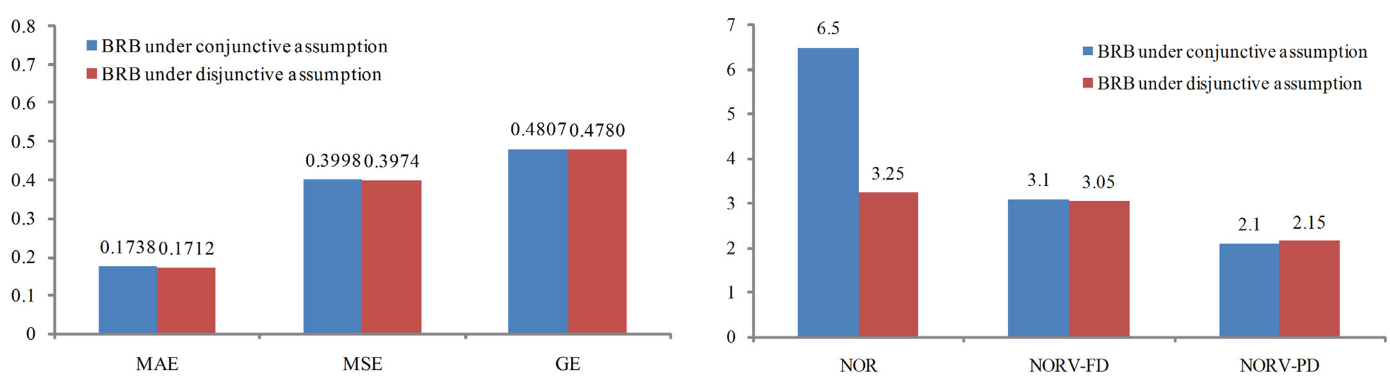

Fig. 16. Comparison of BRB under conjunctive and disjunctive assumptions.

Table 9
Comparison of ten BRBs.

No.	Year	Description	assumption	GE	MAE	Size (test)	No. of parameters	NOR
1	2007 [17]	Local training	conjunctive	0.6724	0.2223	2008	353	56
2	2009 [18]	Online updating	conjunctive	0.8455	0.3954	2008	353	56
3	2010 [35]	Sequential learning	conjunctive	1.2044	0.8506	17	34	5
4	2011 [31]	Adaptive training	conjunctive	0.6565	0.2064	2008	353	56
5	2013 [50]	Extended BRB	conjunctive	12.4361	0.2169	2008	18,902	900
6	2015 [51]	Approximate causal inference	conjunctive	0.6515	0.2014	2008	92	15
7	2016 [34]	Dynamic rule adjustment	conjunctive	0.5113	0.2080	2008	43	6
8	2017 [43]	Bi-level BRB	disjunctive	0.5779	0.1941	2008	40	5
9	**This study**	**JOPS algorithm**	**conjunctive**	**0.4807**	**0.1738**	**2008**	**43**	**6**
10	**This study**	**JOPS algorithm**	**disjunctive**	**0.4780**	**0.1712**	**2008**	**22**	**3**

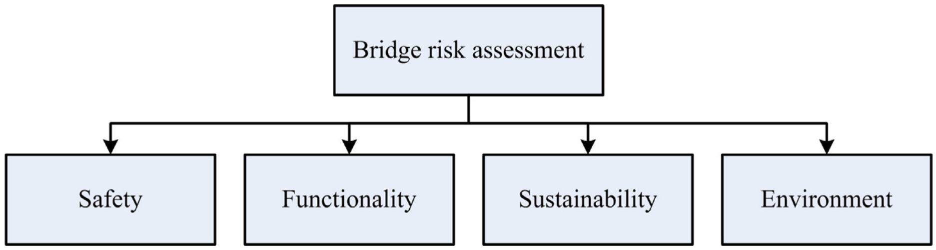

Fig. 17. Hierarchical structure for bridge risk assessment.

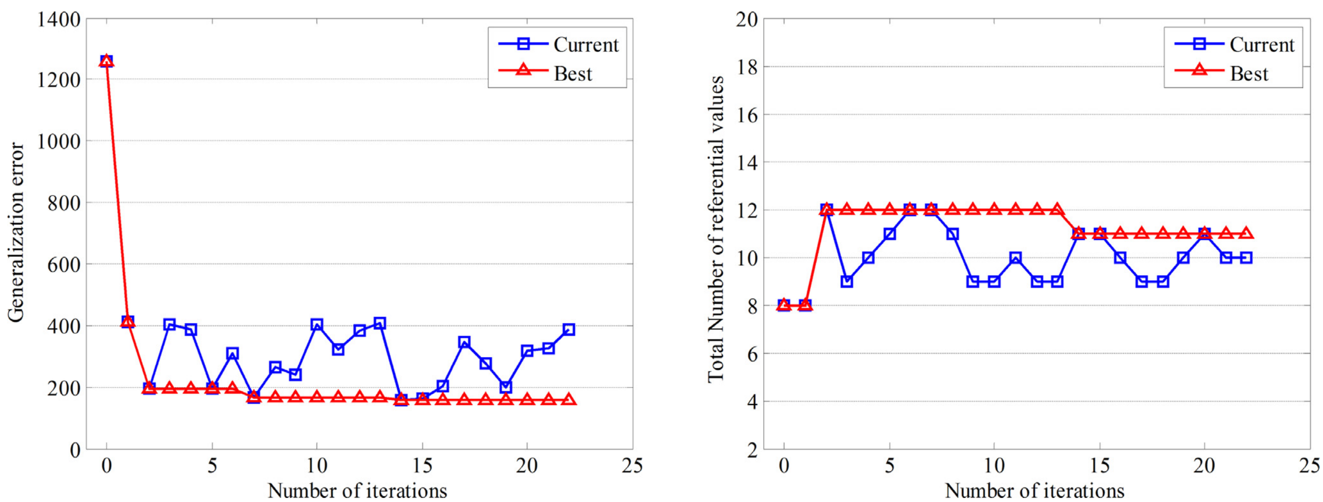

Fig. 18. Changes of the generalization error and the number of referential values with the 22 iterations.

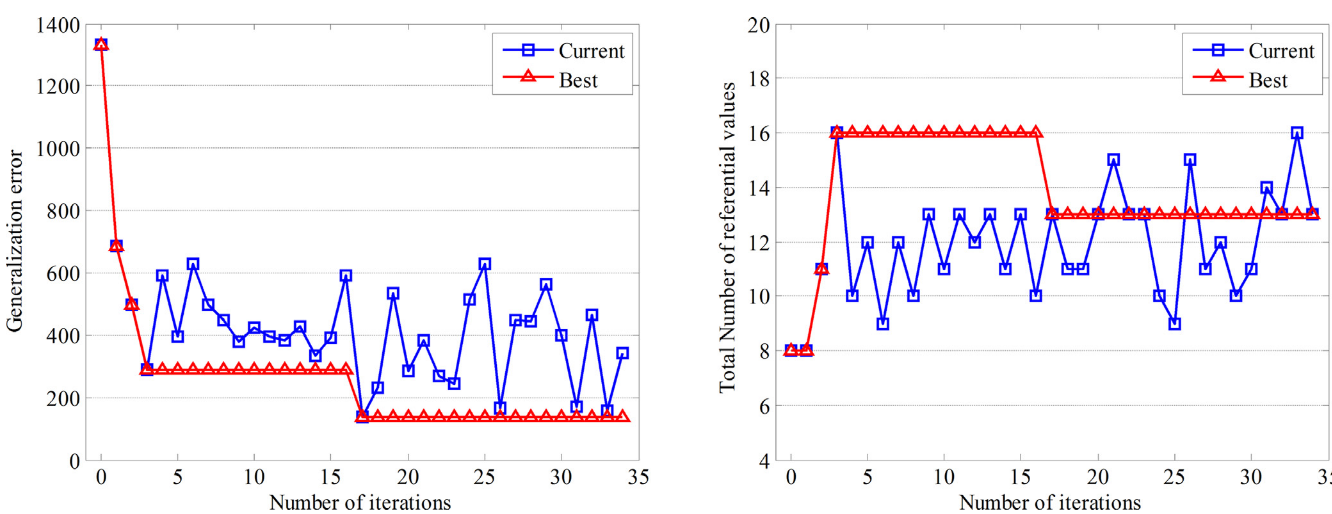

Fig. 19. Changes of the generalization error and the number of referential values with the 34 iterations.

$\varepsilon = 0.1$ and $S_1 = 500$ for running the PODE algorithm, and $S_2 = 3$ for running the JOPS algorithm.

5.2.2. BRB optimization and comparative analysis

For the BRB under conjunctive assumption, Fig. 18 shows the total number of referential values and the generalization error change with the iterative progress in the proposed method. Finally, these values are 0.9836 and 11, respectively, and ⟨0.9836, (4. 2, 3, 2)⟩ is regarded as the optimal parameters and structure of the BRB, which is listed at Table D1 shown in Appendix D.

For the BRB under disjunctive assumption, as shown in Fig. 19, the total number of referential values and the generalization error are reached to a fixed final value, namely 13 and 0.9941, respectively, while the algorithm is stopped. ⟨0.9941, (4, 3, 4, 2)⟩ is selected as the optimal parameters and structure of the BRB, which is listed at Table D2 shown in Appendix D.

To further verify the validity of the proposed method, seven decision models developed by the ANN, ERL, and MRA are used in bridge risk assessment, named as BP-ANN, ER1, ER2, MRA3, MRA7, MRA8, and MRA9 based on [38], where ER1, MRA3, and MRA8 all use zero to represent the minimum risk score of each bridge struc-

Table 10
Comparison of BRBS and other models.

Performance criterion	Models							This study	
	BP-ANN	ER1	MRA3	MRA8	ER2	MRA7	MRA9	**Conjunctive**	**Disjunctive**
MAPE (%)	9.6294	22.4544	18.5775	23.9799	18.7808	24.1941	19.1456	**2.5999**	**6.7620**
RMSE	4.1871	8.9255	10.9527	10.4653	11.2736	10.3510	11.3653	**2.8060**	**3.0916**
R	0.9834	0.9077	0.8687	0.8794	0.8918	0.8796	0.8904	**0.9910**	**0.9911**

ture, and ER2, MRA7 and MRA9 all take no account of zero to stand for the minimum risk score. By comparing the results of these decision models with the BRB under conjunctive and disjunctive assumptions in Table 10, it is validated that the proposed joint optimization method can help the BRBS to construct the optimal parameters and structure of the BRB and produce satisfactory results in terms of different criteria, where the BRB under conjunctive assumption achieves the best results in the MAPE, RMSE, and R, which are slightly better than the BRB under disjunctive assumption and significantly better than the other decision models.

In summary, the comparison results have shown that the proposed joint optimization method can optimize the BRB in both conjunctive and disjunctive assumptions and provide superior inference performances, with the lower MAPE, RMSE, and higher R than other studies in bridge risk assessment.

6. Conclusion and perspectives

This present work mainly contributed to a joint optimization method implemented by the JOPS algorithm which actually consists of two main algorithms, namely the SOHS algorithm and the PODE algorithm. The former can obtain the structure of a new BRB using a heuristic strategy and the latter is applied to obtain the optimized parameters of the BRB based on the DE algorithm. The numerical study in oil pipeline leak detection and bridge risk assessment demonstrated that the proposed method could effectively optimize the BRBS to achieve the compact structure for a BRB. Some other contributions can be summarized into three aspects below:

(1) The previous studies of the BRBS have shown that it is important and necessary to optimize the parameters and structure of a BRB together. However, the training error is usually used to determine the optimal structure of a BRB after optimizing parameters. Thus, by using the Hoeffding inequality theorem, the generalization error using as a new criterion was proposed and justified in this paper to determine the optimal parameters and structure of a BRB.
(2) A heuristic strategy was proposed to efficiently optimize the structure of a BRB, and it can avoid traversing all possible BRBs for obtaining the optimal structure of a BRB. Moreover, based on the DE algorithm, a generic technique is proposed to optimize the parameters of a BRB in different parameter optimization models. Both of them are essential parts of the joint optimization method on parameter and structure.
(3) In the numerical studies of oil pipeline detection and bridge risk assessment, on the one hand, the optimization results of the BRB under conjunctive and disjunctive assumptions clearly showed the generality of the joint optimization method; On the another hand, the comparative results illustrated that the joint optimization method not only could optimize the parameters and structure of a BRB simultaneously, but also could provide superior performance in the terms of accuracy and number of parameters involved.

For the future research, the combinatorial explosion problem, which is caused by too many antecedent attributes or the corresponding referential values, is one of challenges to be addressed to prevent the application of the BRBS in many decision problems.

Acknowledgments

This research was supported by the National Natural Science Foundation of China (Nos. 61773123, 71371053, 71501047 and 71701050), the Humanities and Social Science Foundation of the Ministry of Education under Grant (No. 14YJC630056), the Natural Science Foundation of Fujian Province, China (No. 2015J01248).

Appendix A. ER-based inference method of BRBS

The ER-based inference method of the BRBS mainly includes three steps: (A.1) calculation of the individual matching degree of the given input; (A.2) calculation of the activated weight for each belief rule; and (A.3) integration of the activated belief rule using the ER algorithm. More details can be found in [13].

A.1. Calculation of the individual matching degree of the given input

Suppose $\boldsymbol{x}=(x_1, x_2,\ldots, x_M)$ is the input vector, x_i represents the input data of the ith antecedent attribute, then the individual matching degree can be transformed using rule- or utility-based equivalence transformation techniques [48].

$$\alpha_{i,j} = \frac{u(A_{i,j+1}) - x_i}{u(A_{i,j+1}) - u(A_{i,j})} \text{ and}$$
$$\alpha_{i,j+1} = 1 - \alpha_{i,j}, \text{ if } u(A_{i,j}) \le x_i \le u(A_{i,j+1}), \tag{A1}$$

$$\alpha_{i,k} = 0, \text{ for } k = 1,\ldots,J_i \text{ and } k \ne j, j+1, \tag{A2}$$

where $A_{i,j}$ represents the jth referential value for the ith antecedent attribute, $u(A_{i,j})$ represents the utility value of $A_{i,j}$, $\alpha_{i,j}$ represents the individual matching degree of the given input x_i to $A_{i,j}$.

Finally, the distribution of the individual matching degree is represented as follows:

$$S(x_i) = \left\{\left(A_{i,j}, \alpha_{i,j}\right); i = 1,\ldots,M; j = 1,\ldots,J_i\right\}, \tag{A3}$$

where M is the number of antecedent attributes in the BRB, and J_i is the number of referential values for the ith antecedent attribute.

A.2. Calculation of the activated weight for each belief rule

While the BRB is constructed under the conjunctive assumption, the activated weight for the kth rule is calculated as follows:

$$w_k = \frac{\theta_k \prod_{i=1}^{M} (\alpha_i^k)^{\bar{\delta}_i}}{\sum_{l=1}^{L} \left(\theta_l \prod_{i=1}^{M} \left(\alpha_i^l\right)^{\bar{\delta}_i}\right)}. \tag{A4}$$

Otherwise, for the BRB under disjunctive assumption, the activated weight of the kth rule is calculated as follows:

$$w_k = \frac{\theta_k \sum_{i=1}^{M} (\alpha_i^k)^{\bar{\delta}_i}}{\sum_{l=1}^{L} \left(\theta_l \sum_{i=1}^{M} (\alpha_i^l)^{\bar{\delta}_i}\right)}, \tag{A5}$$

where α_i^k is the individual matching degree to the ith antecedent attribute in the kth rule, and

$$\bar{\delta}_i = \frac{\delta_i}{\max_{i=1,\ldots,M}\{\delta_i\}}, \tag{A6}$$

where θ_k is the rule weight of the kth rule, δ_i is the attribute weight of the ith antecedent attribute.

A.3. Integration of the activated belief rule using the ER algorithm

Using the ER algorithm [52–54], the combined belief degree β_i can be expressed in the follows:

$$\beta_i = \frac{\prod_{k=1}^{L}\left(w_k\beta_{i,k} + 1 - w_k\sum_{n=1}^{N}\beta_{n,k}\right) - \prod_{k=1}^{L}\left(1 - w_k\sum_{n=1}^{N}\beta_{n,k}\right)}{\sum_{n=1}^{N}\prod_{k=1}^{L}\left(w_k\beta_{n,k} + 1 - w_k\sum_{j=1}^{N}\beta_{j,k}\right) - (N-1)\prod_{k=1}^{L}\left(1 - w_k\sum_{j=1}^{N}\beta_{j,k}\right) - \prod_{k=1}^{L}(1-w_k)}. \tag{A7}$$

Assuming that the ith utility value of the referential value D_i is $u(D_i)$, we calculate the numerical inference output as follows:

$$f(\boldsymbol{x}) = \sum_{i=1}^{N}(u(D_i)\beta_i) + \frac{(u(D_1)+u(D_N))}{2}\left(1 - \sum_{i=1}^{N}\beta_i\right). \tag{A8}$$

Appendix B. Inference output of BRBS-1

Assuming that inference output of BRBS-1 is $f(x)$, the *MSE* of the BRBS-1 for all training data is calculated as follows:

$$MSE = \sum_{i=1}^{6}\left(f(x) - x_i^2\right)^2. \tag{B1}$$

The first order derivative of *MSE* with respect to the inference output $f(x)$ can be represented as follows:

$$\frac{\partial MSE}{\partial f(x)} = 2\sum_{i=1}^{6}\left(f(x) - x_i^2\right) = 12f(x) - 2\sum_{i=1}^{6}x_i^2 = 12f(x) - 4.4. \tag{B2}$$

By supposing the first order derivative being zero, we can get the inference output that is able to ensure the minimal error and it is represented as follows:

$$12f(x) - 4.4 = 0 \Leftrightarrow f(x) = 0.3667. \tag{B3}$$

Appendix C. Proof of Theorem 2

Proof. For each input-output data pair ($\boldsymbol{x_i}$, y_i) ($i=1, 2, \ldots, T$), the absolute error between the inference output of BRBS and the actual output data can be expressed as follows:

$$R(f(\boldsymbol{x_i})) = |y_i - f(\boldsymbol{x_i})|. \tag{C1}$$

According to Lemma 1 and the boundedness property [47], it is obvious that the error $R(f(\boldsymbol{x_i}))$ can also be treated as a random variable and all of these random variables are independent due to the independence of input data. And the bounded interval of $R(f(\boldsymbol{x_i}))$ is expressed as follows:

$$0 \le R(f(\boldsymbol{x_i})) \le |u(D_N) - u(D_1)|. \tag{C2}$$

The mean error of these $R(f(\boldsymbol{x_i}))$ is then expressed as follows:

$$R(f) = \frac{R(f(\boldsymbol{x_1})) + \cdots + R(f(\boldsymbol{x_T}))}{T}, \tag{C3}$$

According to the Hoeffding inequality theorem [35], for $t > 0$, we can obtain the following inequality:

$$\begin{aligned} &P((-R(f)) - (-ER(f)) \ge t) \le \exp\left(-\frac{2Tt^2}{(-|u(D_N)-u(D_1)|)^2}\right) \\ \Leftrightarrow\ &P(ER(f) - R(f) \ge t) \le \exp\left(-\frac{2Tt^2}{(u(D_N)-u(D_1))^2}\right) \end{aligned}. \tag{C4}$$

Note that $F(\boldsymbol{x}) = \{f_1(\boldsymbol{x}), \ldots, f_s(\boldsymbol{x})\}$ and $S = 2^d$, we can obtain

$$\begin{aligned} P(\exists f \in F : ER(f) - R(f) \ge t) &= P\left(\bigcup_{f\in F}\{ER(f) - R(f) \ge t\}\right) \\ &\le \sum_{f\in F} P(ER(f) - R(f) \ge t) \\ &\le 2^d \exp\left(-\frac{2Tt^2}{(u(D_N)-u(D_1))^2}\right) \end{aligned}. \tag{C5}$$

Next, we can obtain

$$P(ER(f) - R(f) < t) \le 1 - 2^d \exp\left(-\frac{2Tt^2}{(u(D_N) - u(D_1))^2}\right). \tag{C6}$$

By assuming

$$r = 2^d \exp\left(-\frac{2Tt^2}{(u(D_N) - u(D_1))^2}\right). \tag{C7}$$

And then to replace the variable t, we can obtain

$$P\left(ER(f) < R(f) + \sqrt{\frac{(d\ln 2 - \ln r)(u(D_N) - u(D_1))^2}{2T}}\right) \le 1 - r. \tag{C8}$$

Finally, we can obtain the following inequality:

$$ER(f) < R(f) + \sqrt{\frac{(d\ln 2 - \ln r)(u(D_N) - u(D_1))^2}{2T}}. \tag{C9}$$

Appendix D. Trained BRB under conjunctive and disjunctive assumptions for bridge risk assessment

Table D1
Trained BRB under conjunctive assumption.

Rule no.	Rule weight	Safety $\delta_1 = 0.5235$	Functionality $\delta_2 = 0.9998$	Sustainability $\delta_3 = 0.6502$	Environment $\delta_4 = 0.4834$	Risk scores
1	0.4925	−1.0000	−1.0000	−1.0000	−1.0000	{0.1185, 0.0347, 0.1964, 0.1542, 0.4962}
2	0.2816	−1.0000	−1.0000	−1.0000	4.0000	{0.0895, 0.2194, 0.0925, 0.3108, 0.2877}
3	0.2670	−1.0000	−1.0000	2.0341	−1.0000	{0.0692, 0.2181, 0.1073, 0.4211, 0.1843}
4	0.3784	−1.0000	−1.0000	2.0341	4.0000	{0.1032, 0.0119, 0.2361, 0.3005, 0.3483}
5	0.5933	−1.0000	−1.0000	4.0000	−1.0000	{0.1463, 0.1844, 0.0104, 0.3729, 0.2860}
6	0.5681	−1.0000	−1.0000	4.0000	4.0000	{0.0161, 0.2202, 0.2687, 0.1583, 0.3366}
7	0.2614	−1.0000	4.0000	−1.0000	−1.0000	{0.1648, 0.1016, 0.2937, 0.3727, 0.0671}
8	0.4147	−1.0000	4.0000	−1.0000	4.0000	{0.4253, 0.0187, 0.0818, 0.1439, 0.3304}
9	0.5349	−1.0000	4.0000	2.0341	−1.0000	{0.3385, 0.2213, 0.1930, 0.1553, 0.0919}
10	0.3408	−1.0000	4.0000	2.0341	4.0000	{0.0731, 0.2583, 0.2625, 0.1744, 0.2317}
11	0.3470	−1.0000	4.0000	4.0000	−1.0000	{0.2052, 0.1610, 0.4319, 0.1621, 0.0398}
12	0.8989	−1.0000	4.0000	4.0000	4.0000	{0.1062, 0.4448, 0.1136, 0.2632, 0.0721}
13	0.6949	0.0000	−1.0000	−1.0000	−1.0000	{0.6900, 0.3099, 0.0000, 0.0001, 0.0000}
14	0.0001	0.0000	−1.0000	−1.0000	4.0000	{0.2687, 0.2856, 0.0898, 0.0952, 0.2607}
15	0.0590	0.0000	−1.0000	2.0341	−1.0000	{0.4974, 0.0059, 0.0874, 0.0479, 0.3613}
16	0.1352	0.0000	−1.0000	2.0341	4.0000	{0.6338, 0.2932, 0.0726, 0.0003, 0.0001}
17	0.8058	0.0000	−1.0000	4.0000	−1.0000	{0.1801, 0.0413, 0.2070, 0.3908, 0.1808}
18	0.7147	0.0000	−1.0000	4.0000	4.0000	{0.0829, 0.0863, 0.1780, 0.2464, 0.4064}
19	0.0000	0.0000	4.0000	−1.0000	−1.0000	{0.4452, 0.0135, 0.2606, 0.1341, 0.1465}
20	0.0001	0.0000	4.0000	−1.0000	4.0000	{0.0875, 0.4077, 0.1147, 0.1421, 0.2480}
21	0.1664	0.0000	4.0000	2.0341	−1.0000	{0.8611, 0.0002, 0.1026, 0.0213, 0.0148}
22	0.8171	0.0000	4.0000	2.0341	4.0000	{0.2839, 0.0001, 0.5259, 0.1498, 0.0403}
23	0.4121	0.0000	4.0000	4.0000	−1.0000	{0.0017, 0.0720, 0.1461, 0.1144, 0.6658}
24	0.9998	0.0000	4.0000	4.0000	4.0000	{0.0006, 0.0010, 0.0012, 0.3815, 0.6158}
25	0.8674	3.0000	−1.0000	−1.0000	−1.0000	{0.0134, 0.1391, 0.2788, 0.1553, 0.4134}
26	0.2936	3.0000	−1.0000	−1.0000	4.0000	{0.0008, 0.1602, 0.3113, 0.2302, 0.2975}
27	0.0413	3.0000	−1.0000	2.0341	−1.0000	{0.0004, 0.0001, 0.0005, 0.3302, 0.6687}
28	0.0000	3.0000	−1.0000	2.0341	4.0000	{0.1168, 0.1055, 0.4005, 0.0438, 0.3333}
29	0.0196	3.0000	−1.0000	4.0000	−1.0000	{0.0744, 0.0945, 0.1817, 0.3517, 0.2977}
30	0.1260	3.0000	−1.0000	4.0000	4.0000	{0.0000, 0.0005, 0.0002, 0.4881, 0.5112}
31	0.6644	3.0000	4.0000	−1.0000	−1.0000	{0.0421, 0.3583, 0.1647, 0.2702, 0.1646}
32	0.7457	3.0000	4.0000	−1.0000	4.0000	{0.0271, 0.3229, 0.2902, 0.2127, 0.1472}
33	0.0649	3.0000	4.0000	2.0341	−1.0000	{0.0502, 0.0128, 0.0399, 0.0202, 0.8769}
34	0.4781	3.0000	4.0000	2.0341	4.0000	{0.0000, 0.0000, 0.0232, 0.0000, 0.9767}
35	0.0375	3.0000	4.0000	4.0000	−1.0000	{0.0966, 0.1936, 0.3950, 0.2805, 0.0344}
36	0.0000	3.0000	4.0000	4.0000	4.0000	{0.2690, 0.1066, 0.2065, 0.3610, 0.0570}
37	0.7570	4.0000	−1.0000	−1.0000	−1.0000	{0.2772, 0.0046, 0.1362, 0.1812, 0.4009}
38	0.3607	4.0000	−1.0000	−1.0000	4.0000	{0.2408, 0.0052, 0.1030, 0.1660, 0.4851}
39	0.2866	4.0000	−1.0000	2.0341	−1.0000	{0.2357, 0.0304, 0.1687, 0.4945, 0.0708}
40	0.5887	4.0000	−1.0000	2.0341	4.0000	{0.0199, 0.3283, 0.2958, 0.2679, 0.0881}
41	0.3256	4.0000	−1.0000	4.0000	−1.0000	{0.1059, 0.4580, 0.1009, 0.2688, 0.0664}
42	0.6694	4.0000	−1.0000	4.0000	4.0000	{0.1315, 0.2040, 0.3054, 0.2179, 0.1412}
43	0.3855	4.0000	4.0000	−1.0000	−1.0000	{0.1369, 0.1642, 0.3251, 0.2740, 0.0997}
44	0.3471	4.0000	4.0000	−1.0000	4.0000	{0.1111, 0.1141, 0.2645, 0.0849, 0.4253}
45	0.4770	4.0000	4.0000	2.0341	−1.0000	{0.5918, 0.0323, 0.2038, 0.1384, 0.0336}
46	0.6951	4.0000	4.0000	2.0341	4.0000	{0.1337, 0.4535, 0.1017, 0.1164, 0.1946}
47	0.1004	4.0000	4.0000	4.0000	−1.0000	{0.0926, 0.3711, 0.0979, 0.2759, 0.1625}
48	0.5791	4.0000	4.0000	4.0000	4.0000	{0.3149, 0.0688, 0.0268, 0.2965, 0.2931}

Table D2
Trained BRB under disjunctive assumption.

Rule no.	Rule weight	Safety $\delta_1 = 0.3472$	Functionality $\delta_2 = 0.2258$	Sustainability $\delta_3 = 0.8786$	Environment $\delta_4 = 0.5058$	Risk scores
1	0.0343	−1.0000	2.4726	2.8379	4.0000	{0.0000, 0.1916, 0.3394, 0.4518, 0.0172}
2	0.0385	0.4471	−1.0000	0.1622	−1.0000	{0.9999, 0.0000, 0.0000, 0.0000, 0.0000}
3	0.9027	4.0000	4.0000	4.0000	–	{0.0059, 0.0011, 0.0008, 0.0000, 0.9922}
4	0.2369	2.0236	–	–	–	{0.2825, 0.0086, 0.0006, 0.1713, 0.5370}
5	0.0001	–	–	−1.0000	–	{0.5147, 0.1004, 0.1508, 0.2220, 0.0120}

References

[1] R.L. Dou, C. Zong, M.Q. Li, An interactive genetic algorithm with the interval arithmetic based on hesitation and its application to achieve customer collaborative product configuration design, Appl. Soft Comput. 38 (2016) 384–394.

[2] R.L. Dou, Y.B. Zhang, G.F. Nan, Application of combined Kano model and interactive genetic algorithm for product customization, J. Intell. Manuf. (2016) 1–16.

[3] R.L. Dou, C. Zong, G.F. Nan, Multi-stage interactive genetic algorithm for collaborative product customization, Knowl.-Based Syst. 92 (2016) 43–54.

[4] M.S. Hossain, F. Ahmed, K.Andersson Fatema-Tuj-Johora, A belief rule base expert system to assess tuberculosis under uncertainty, J. Med. Syst. 41 (43) (2017) 1–11.

[5] M.S. Hossain, S. Rahaman, R. Mustafa, K. Andersson, A belief rule-based expert system to assess suspicion of acute coronary syndrome (ACS) under uncertainty, Soft Comput. 6 (2017) 1–16.

[6] R. Sun, Robust reasoning: integrating rule-based and similarity-based reasoning, Artif. Intell. 75 (2) (1995) 241–295.

[7] L.A. Zadeh, Fuzzy sets, Inf. Control 8 (69) (1965) 338–353.

[8] Y.W. Chen, H.S. Poon, J.B. Yang, D.L. Xu, D. Zhang, S. Acomb, Belief rule-based system for portfolio optimisation with nonlinear cash-flows and constraints, Eur. J. Oper. Res. 223 (3) (2012) 775–784.

[9] G.L. Kong, D.L. Xu, R. Body, J.B. Yang, K. Machway-Jones, S. Carley, A belief rule-based decision support system for clinical risk assessment of cardiac chest pain, Eur. J. Oper. Res. 219 (3) (2012) 564–573.

[10] B. Li, H.W. Wang, J.B. Yang, M. Guo, C. Qi, A belief-rule-based inventory control method under nonstationary and uncertain demand, Expert Syst. Appl. 38 (12) (2011) 14997–15008.
[11] J.W. Hall, D.I. Blockley, J.P. Davis, Uncertain inference using interval probability theory, Int. J. Approximate Reasoning 19 (3-4) (1998) 247–264.
[12] W. Cui, D.I. Blockley, Interval probability theory for evidential support, Int. J. Intell. Syst. 5 (2) (1990) 183–192.
[13] J.B. Yang, J. Liu, J. Wang, H.S. Sii, H.W. Wang, Belief rule-base inference methodology using the evidential reasoning approach-RIMER, IEEE Trans. Syst., Man, Cybern.-Part A 37 (4) (2006) 569–585.
[14] Y.M. Wang, J.B. Yang, D.L. Xu, K.S. Chin, Consumer preference prediction by using a hybrid evidential reasoning and belief rule-based methodology, Expert Syst. Appl. 36 (2009) 8421–8430.
[15] J.B. Yang, Y.M. Wang, D.L. Xu, K.S. Chin, L. Chatton, Belief rule-based methodology for mapping consumer preferences and setting product targets, Expert Syst. Appl. 39 (2012) 4749–4759.
[16] Y. Yang, C. Fu, Y.W. Chen, D.L. Xu, S.L. Yang, A belief rule based expert system for predicting consumer preference in new product development, Knowl.-Based Syst. 94 (2016) 105–113.
[17] D.L. Xu, J. Liu, J.B. Yang, G.P. Liu, J. Wang, L. Jenkinson, R. Jun, Inference and learning methodology of belief-rule-based expert system for pipeline leak detection, Expert Syst. Appl. 32 (2007) 103–113.
[18] Z.J. Zhou, C.H. Hu, J.B. Yang, D.L. Xu, D.H. Zhou, Online updating belief-rule-based system for pipeline leak detection under expert intervention, Expert Syst. Appl. 36 (4) (2009) 7700–7709.
[19] L. Dymova, P. Sevastjanov, K. Kaczmarek, A Forex trading expert system based on a new approach to the rule-base evidential reasoning, Expert Syst. Appl. 51 (2016) 1–13.
[20] Z.J. Zhou, L.L. Chang, C.H. Hu, X.X. Han, 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. 46 (11) (2016) 1529–1543.
[21] J. Jiang, Z.J. Zhou, X.X. Han, B.C. Zhang, X.D. Ling, A new BRB based method to establish hidden failure prognosis model by using life data and monitoring observation, Knowl.-Based Syst. 67 (2014) 270–277.
[22] Z.J. Zhou, C.H. Hu, X.X. Han, H.F. He, X.D. Ling, B.C. Zhang, A model for online failure prognosis subject to two failure modes based on belief rule base and semi-quantitative information, Knowl.-Based Syst. 70 (2014) 221–230.
[23] G.L. Li, Z.J. Zhou, C.H. Hu, L.L. Chang, Z.G. Zhou, F.J. Zhao, A new safety assessment model for complex system based on the conditional generalized minimum variance and the belief rule base, Safety Sci. 93 (2017) 108–120.
[24] G.L. Kong, D.L. Xu, J.B. Yang, X.F. Yin, T.B. Wang, B.G. Jiang, Y.H. Hu, Belief rule-based inference for predicting trauma outcome, Knowl.-Based Syst. 95 (2016) 35–44.
[25] Z.G. Zhou, F. Liu, L.C. Jiao, Z.J. Zhou, J.B. Yang, M.G. Gong, A bi-level belief rule based decision support system for diagnosis of lymph node metastasis in gastric cancer, Knowl.-Based Syst. 54 (2013) 128–136.
[26] Z.G. Zhou, F. Liu, L.L. Li, L.C. Jiao, Z.J. Zhou, J.B. Yang, Z.L. Wang, A cooperative belief rule based decision support system for lymph node metastasis diagnosis in gastric cancer, Knowl.-Based Syst. 85 (2015) 62–70.
[27] L.M. Jiao, Q. Pan, T. Denoeux, Y. Liang, X.X. Feng, Belief rule-based classification system: extension of FRBCS in belief functions framework, Inf. Sci. 309 (2015) 26–49.
[28] L.H. Yang, Y.M. Wang, Q. Su, Y.G. Fu, K.S. Chin, Multi-attribute search framework for optimizing extended belief rule-based systems, Inf. Sci. 370-371 (2016) 159–183.
[29] L.H. Yang, Y.M. Wang, Y.X. Lan, L. Chen, Y.G. Fu, A data envelopment analysis (DEA)-based method for rule reduction in extended belief-rule-based systems, Knowl.-Based Syst. 123 (2017) 174–187.
[30] J.B. Yang, J. Liu, D.L. Xu, J. Wang, H.W. Wang, Optimization model for training belief-rule-based systems, IEEE Trans. Syst., Man, Cybern.-Part A 37 (4) (2007) 569–585.
[31] Y.W. Chen, J.B. Yang, D.L. Xu, Z.J. Zhou, D.W. Tang, Inference analysis and adaptive training for belief rule based system, Expert Syst. Appl. 38 (2011) 12845–12860.
[32] L.L. Chang, J.B. Sun, J. Jiang, M.J. Li, Parameter learning for the belief rule base system in the residual life probability prediction of metalized file capacitor, Knowl.-Based Syst. 73 (2015) 69–80.
[33] L.L. Chang, Y. Zhou, J. Jiang, M.J. Li, X.H. Zhang, Structure learning for belief rule base expert system: a comparative study, Knowl.-Based Syst. 39 (2013) 159–172.
[34] Y.M. Wang, L.H. Yang, Y.G. Fu, L.L. Chang, K.S. Chin, Dynamic rule adjustment approach for optimizing belief rule-base expert system, Knowl.-Based Syst. 96 (2016) 40–60.
[35] Z.J. Zhou, C.H. Hu, J.B. Yang, D.L. Xu, M.Y. Chen, D.H. Zhou, A sequential learning algorithm for online constructing belief rule based systems, Expert Syst. Appl. 37 (2) (2010) 1790–1799.
[36] W. Hoeffding, Probability inequalities for sums of bounded random variables, J. Am. Stat. Assoc. 58 (301) (1963) 13–30.
[37] K. Price, R. Storn, J. Lampinen, Differential Evolution-A Practical Approach to Global Optimization, Springer, 2005.
[38] Y.M. Wang, T.M.S. Elhag, A comparison of neural network, evidential reasoning and multiple regression analysis in modelling bridge risks, Expert Syst. Appl. 32 (2) (2007) 336–348.
[39] T.M.S. Elhag, Y.M. Wang, Risk assessment for bridge maintenance projects: neural networks versus regression techniques, J. Comput. Civil Eng. 21 (6) (2007) 402–409.
[40] L.L. Chang, Z.J. Zhou, Y. You, L.H. Yang, Z.G. Zhou, Belief rule based expert system for classification problems with new rule activation and weight calculation procedures, Inf. Sci. 336 (2016) 75–91.
[41] L.H. Yang, Y.M. Wang, L.L. Chang, Y.G. Fu, A disjunctive belief rule-based expert system for bridge risk assessment with dynamic parameter optimization model, Comput. Ind. Eng. 113 (2017) 459–474.
[42] Z.J. Zhou, C.H. Hu, J.B. Yang, D.L. Xu, D.H. Zhou, Online updating belief-rule-base using the RIMER approach, IEEE Trans. Syst. Man Cybern.-Part A 41 (6) (2011) 1225–1243.
[43] L.L. Chang, Z.J. Zhou, T.J. Liao, Y.W. Chen, H. Yu, L.H. Yang, Belief rule base structure and parameter joint learning under disjunctive assumption for nonlinear complex system modeling, IEEE Trans. Syst., Man, Cybern.-Part A PP (99) (2017) 1–13.
[44] A.R. Barron, Approximation and estimation bounds for artificial neural networks, Mach. Learn. 14 (1994) 115–133.
[45] M. Seeger, PAC-Bayesian generalization error bounds for Gaussian process classification, J. Mach. Learn. Res. 3 (2002) 233–269.
[46] V. Vapnik, O. Chapelle, Bounds on error expectation for support vector machines, Neural Comput. 12 (9) (2000) 2013–2036.
[47] Y.W. Chen, J.B. Yang, D.L. Xu, S.L. Yang, On the inference and approximation properties of belief rule based systems, Inf. Sci. 234 (2013) 121–135.
[48] J.B. Yang, Rule and utility based evidential reasoning approach for multi-attribute decision analysis under uncertainties, Eur. J. Oper. Res. 131 (1) (2001) 31–61.
[49] E. Zitzler, L. Thiele, Multi-objective evolutionary algorithms: a comparative case study and strength Pareto approach, IEEE Trans. Evol. Comput. 3 (4) (1999) 257–271.
[50] J. Liu, L. Martinez, A. Calzada, H. Wang, A novel belief rule base representation, generation and its inference methodology, Knowl.-Based Syst. 53 (2013) 129–141.
[51] Y. Chen, Y.W. Chen, X.B. Xu, C.C. Pan, J.B. Yang, G.K. Yang, A data-driven approximate causal inference model using the evidential reasoning rule, Knowl.-Based Syst. 88 (2015) 264–272.
[52] Y.M. Wang, J.B. Yang, D.L. Xu, Environmental impact assessment using the evidential reasoning approach, Eur. J. Oper. Res. 174 (3) (2006) 1885–1913.
[53] J.B. Yang, D.L. Xu, On the evidential reasoning algorithm for multiple attribute decision analysis under uncertainty, IEEE Trans. Syst. Man Cybern.-Part A 32 (3) (2002) 289–304.
[54] M.J. Zhang, Y.M. Wang, L.H. Li, S.Q. Chen, A general evidential reasoning algorithm for multi-attribute decision analysis under interval uncertainty, Eur. J. Oper. Res. 257 (3) (2017) 1005–1015.
[55] Y.M. Wang, T.M.S. Elhag, Evidential reasoning approach for bridge condition assessment, Expert Syst. Appl. 34 (1) (2008) 689–699.