



Inference analysis and adaptive training for belief rule based systems

Yu-Wang Chen^{a,*}, Jian-Bo Yang^a, Dong-Ling Xu^a, Zhi-Jie Zhou^b, Da-Wei Tang^a

^a Manchester Business School, The University of Manchester, Manchester M15 6PB, UK

^b High-Tech Institute of Xi'an, Xi'an, Shaanxi 710025, PR China

ARTICLE INFO

Keywords:

Belief rule base
Evidential reasoning
Inference analysis
Adaptive training
Multi-modal function
Leak detection

ABSTRACT

Belief rule base (BRB) systems are an extension of traditional IF-THEN rule based systems and capable of capturing complicated nonlinear causal relationships between antecedent attributes and consequents. In a BRB system, various types of information with uncertainties can be represented using belief structures, and a belief rule is designed with belief degrees embedded in its possible consequents. For a set of inputs to antecedent attributes, inference in BRB is implemented using the evidential reasoning (ER) approach. In this paper, the inference mechanism of the ER algorithm is analyzed first and its patterns of monotonic inference and nonlinear approximation are revealed. For a practical BRB system, it is difficult to determine its parameters accurately by using only experts' subjective knowledge. Moreover, the appropriate adjustment of the parameters of a BRB system using available historical data can lead to significant improvement on its prediction performance. In this paper, a training data selection scheme and an adaptive training method are developed for updating BRB parameters. Finally, numerical studies on a multi-modal function and a practical pipeline leak detection problem are conducted to illustrate the functionality of BRB systems and validate the performance of the adaptive training technique.

© 2011 Elsevier Ltd. All rights reserved.

1. Introduction

Among knowledge based systems, rule base is one of the most common and easily understood schemes for expressing various types of knowledge (Davis, 1986; Hayes-Roth, 1985). Moreover, it has also been convincingly argued that other knowledge representation forms can be transformed into rule-based schemes (Nilsson, 1982; Sun, 1995). As such, rule-based systems, usually constructed from human knowledge in the formats of IF-THEN rules, have seen wide applications in fields of artificial intelligence and decision support systems (Azibi & Vanderpooten, 2002; Ligeza, 2006; Negnevitsky, 2002).

On the basis of the concept of belief structures and the evidential reasoning (ER) approach, Yang, Liu, Wang, Sii, and Wang (2006) proposed a generic belief rule based inference system, which extends the applicability of traditional rule based systems. In a belief rule base (BRB) system, various types of input information with uncertainties are represented using belief structures, and belief rules are designed with an extended IF-THEN scheme, in which each possible consequent is associated with a belief degree. An input to a BRB system is first transformed into a belief distribution, which is then used to calculate the activation weights of belief rules. The output of a BRB system is generated using the ER approach. Compared with conventional rule based systems, a BRB

system is capable of not only incorporating a more informative knowledge representation scheme for both quantitative data and qualitative information with uncertainties, but also capturing more complicated nonlinear causal relationships between antecedent attributes and consequents (Yang et al., 2006).

In the implementation of a BRB system, its belief rule base and inference engine are two essential components. The belief rule base can be initially established by human experts with domain-specific knowledge and facts. The inference engine is used to reason from the belief rule base for generating output from a specific set of inputs to its antecedent attributes. Since the ER algorithm is the kernel reasoning engine of a BRB system, it is necessary to investigate thoroughly and understand properly its inference mechanism. Yang and Xu (2002a) studied the typical non-linear reasoning patterns of the ER approach under harmonic, quasi-harmonic, and contradictory evidence. Moreover, training techniques are developed to tune the parameters of an initial BRB system when observed input–output data are available. Such training is necessary for a large scale BRB system because it is difficult to determine its parameters accurately by using only experts' subjective knowledge. Yang, Liu, Xu, Wang, and Wang (2007) and Zhou et al. (2009, 2010) proposed relevant training schemes for locally updating the parameters of a BRB system.

In this paper, we first provide a more general analysis about the ER inference patterns on both the belief degrees and the activation weights. A heuristic scheme for training data selection is then proposed in light of the characteristics of BRB. Since updating the

* Corresponding author.

E-mail address: yu-wang.chen@mbs.ac.uk (Y.-W. Chen).

referential values of antecedent attributes is crucial for identifying the critical points of a BRB system, an adaptive training method is developed to fine tune all parameters in a BRB system, including rule weights, attribute weights, basic belief degrees and the referential values of antecedent attributes. The optimization problems are solved using the nonlinear optimization solver in the Optimization Toolbox of Matlab. Two numerical studies are provided to illustrate the inference mechanism and validate the adaptive training method. The first one is concerned with a numerical multi-modal function, and the second deals with a practical pipeline leak detection problem.

The rest of this paper is organized as follows: In the following section, the concepts of belief structures and BRB are briefly discussed, followed by a schematic architecture for BRB inference systems. In Section 3, the inference patterns of the ER approach are studied and illustrated, including monotonic inference and nonlinear approximation. The training data selection scheme and the adaptive training method for updating BRB parameters are developed in Section 4. In Section 5, a numerical multi-modal function and a practical application on pipeline leak detection are given to illustrate the functionality of BRB systems and validate the performance of the adaptive training method. The paper is concluded in Section 6.

2. Belief rule based systems

2.1. Belief structure

In BRB systems, belief structures are used to represent various types of information with uncertainties. Generally, a belief structure is characterized by a distribution, which was originally designed to model a subjective assessment with uncertainty (Yang & Singh, 1994). To evaluate the safety of an engineering system, for example, an expert may assert that he is 50% sure that the safety is *average* and 30% sure it is *good*. Here, system safety is an attribute and *average* and *good* are linguistic variables for describing the relative states of system safety, which are referred to as distinctive evaluation grades. The percentage values of 50% and 30% are referred to as the degrees of belief on relevant evaluation grades. In engineering system safety expressions, a set of evaluation grades can be defined as follows (Liu, Yang, Ruan, Martinez, & Wang, 2008),

$$H = \{poor(H_1), fair(H_2), average(H_3), good(H_4)\}.$$

Then, the corresponding belief distribution in the above example is represented as,

$$\alpha = \{0, 0, 0.5, 0.3\}.$$

Note that the assessment information is incomplete as the total belief degree is less than 1, and the missing belief degree of 0.2 represents the degree of ignorance.

Suppose a set of N_i distinctive evaluation grades can provide complete standards for evaluating an attribute x_i ($i = 1, \dots, T$), as represented by,

$$H(x_i) = \{H_{i,n}, n = 1, \dots, N_i\},$$

where $H_{i,n}$ denotes the n th evaluation grade for attribute x_i . Accordingly, the belief distribution of attribute x_i can be formulated as follows:

$$S(x_i) = \{(H_{i,n}, \alpha_{i,n}), n = 1, \dots, N_i\},$$

where $\alpha_{i,n}$ is the belief degree to which x_i is assessed to evaluation grade $H_{i,n}$, and $\alpha_{i,n} \geq 0$ and $\sum_{n=1}^{N_i} \alpha_{i,n} \leq 1$. The belief information is regarded as complete if $\sum_{n=1}^{N_i} \alpha_{i,n} = 1$ and incomplete if $\sum_{n=1}^{N_i} \alpha_{i,n} < 1$. Note that both quantitative data and qualitative infor-

mation, with or without uncertainties, can be transformed into the united belief distribution.

2.2. Belief rule base

Over the past few decades, IF-THEN rules have been frequently used to construct knowledge based systems. More recently, Yang et al. (2006) proposed a new knowledge representation scheme by building a generic IF-THEN rule, referred to as belief rule, using the above mentioned belief structure. In a belief rule, all possible consequents are associated with belief degrees, and the weights of both antecedent attributes and rules are also evaluated. Such a belief rule base is capable of capturing vagueness, incompleteness, and nonlinear causal relationships between antecedent attributes and consequents, and traditional IF-THEN rules can be represented as a special case (Yang et al., 2006). Formally, a belief rule can be defined as follows:

$$\begin{aligned} & \text{IF } x_1 \text{ is } A_1^k \wedge x_2 \text{ is } A_2^k \wedge \dots \wedge x_{T_k} \text{ is } A_{T_k}^k, \\ R_k : & \text{ THEN } \{(D_1, \beta_{1,k}), (D_2, \beta_{2,k}), \dots, (D_N, \beta_{N,k})\}, \left(\sum_{n=1}^N \beta_{n,k} \leq 1 \right), \\ & \text{with rule weight } \theta_k, \\ & \text{and attribute weight } \delta_{1,k}, \delta_{2,k}, \dots, \delta_{T_k,k}, k \in \{1, \dots, L\}. \end{aligned} \quad (1)$$

where x_1, x_2, \dots, x_{T_k} denote the antecedent attributes in the k th rule. These attributes belong to the whole set of antecedent attributes $X = \{x_i; i = 1, \dots, T\}$, in which each element takes values (or propositions) from an array of finite sets $A = \{A_1, \dots, A_T\}$. The vector $A_i = \{A_{i,n}; n = 1, \dots, N_i = |A_i|\}$ is defined as the set of referential values for antecedent attribute x_i . In the k th rule, A_i^k represents the referential value taken by the i th antecedent attribute. T_k denotes the total number of antecedent attributes in the k th rule. $\beta_{n,k}$ represents the belief degree to which D_n is believed to be the consequent, given the logical relationship of the k th rule $F_k: x_1 \text{ is } A_1^k \wedge x_2 \text{ is } A_2^k \wedge \dots \wedge x_{T_k} \text{ is } A_{T_k}^k$. If $\sum_{n=1}^N \beta_{n,k} = 1$, the k th rule is said to be complete; otherwise, it is incomplete. The extreme case $\sum_{n=1}^N \beta_{n,k} = 0$ denotes total ignorance on the consequent. Note that the element D_n in the set of consequents $D = \{D_n; n = 1, \dots, N\}$ can either be a conclusion or an action (Yang et al., 2006).

As defined above, a belief IF-THEN rule represents a functional mapping between antecedents and consequents with uncertainties, and it can provide a more informative and realistic scheme than traditional IF-THEN rules. Furthermore, the parameters, including belief degrees $\beta_{n,k}$, rule weights θ_k and attribute weights $\delta_{i,k}$ can be assigned initially by experts and subsequently trained or updated using appropriate learning algorithms if data regarding the inputs and outputs of BRB systems are available.

Once a generic BRB $R = \langle X, A, D, F \rangle$ is established, the knowledge embedded in these belief rules can be used to perform inference for a specific input vector. The inference process of BRB systems is briefly discussed in the following subsection.

2.3. Inference of BRB systems

To a multiple input BRB system, suppose the following T attributes are all the factors influencing the system's output, i.e.,

$$X = \{x_i, i = 1, \dots, T\}.$$

The relative weights $\delta = \{\delta_1, \dots, \delta_T\}$ of the T attributes can be initially estimated using an appropriate method such as the eigenvector or geometric mean method. The weight δ_i represents the relative important of attribute x_i and is usually normalized, so that,

$$0 \leq \delta_i \leq 1 \quad \text{and} \quad \sum_{i=1}^T \delta_i = 1.$$

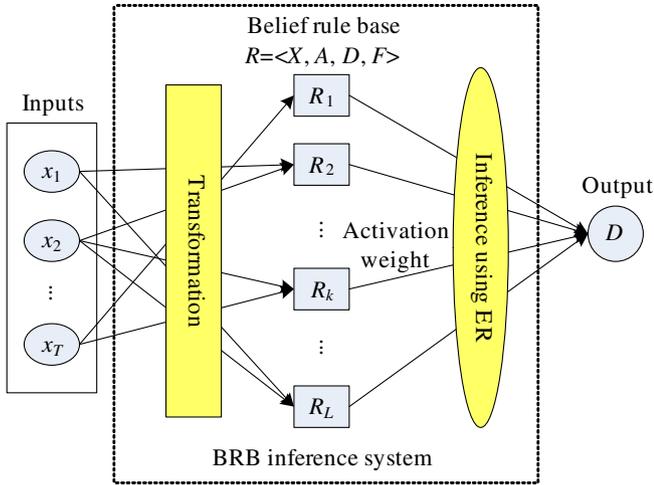


Fig. 1. Single-layer BRB inference system.

Then, the inference of estimated output on a set of inputs can be summarized as three main steps: (1) input transformation, which transforms various formats of information into belief distributions, (2) calculation of activation weights, which matches a specific in-

$$\beta_n = \frac{(w_k \beta_{n,k} + 1 - w_k \sum_{i=1}^N \beta_{i,k})(w_l \beta_{n,l} + 1 - w_l \sum_{i=1}^N \beta_{i,l}) - (1 - w_k \sum_{i=1}^N \beta_{i,k})(1 - w_l \sum_{i=1}^N \beta_{i,l})}{\sum_{j=1}^N [(w_k \beta_{j,k} + 1 - w_k \sum_{i=1}^N \beta_{i,k})(w_l \beta_{j,l} + 1 - w_l \sum_{i=1}^N \beta_{i,l}) - (1 - w_k \sum_{i=1}^N \beta_{i,k})(1 - w_l \sum_{i=1}^N \beta_{i,l})]} \quad (2)$$

put vector to belief rules, and (3) inference using evidential reasoning. Here we consider a basic single-layer multiple input single output (MISO) system, whose inference architecture is illustrated in Fig. 1.

As shown in Fig. 1, various types of input information are first transformed into the belief structure using the rule and utility based techniques (Yang, 2001), so that both quantitative data and qualitative information with uncertainties can be handled in a consistent manner. To simplify the discussion and focus on the main research issues of this paper, suppose all inputs are numerical

$$\beta_n = \frac{w_k(w_l \beta_{n,l} + 1 - w_l) \beta_{n,k} + (1 - w_k) w_l \beta_{n,l}}{w_k(w_l \beta_{n,l} + 1 - w_l) \beta_{n,k} + (1 - w_k) w_l \beta_{n,l} + \sum_{j=1, j \neq n}^N (w_k(w_l \beta_{j,l} + 1 - w_l) \beta_{j,k} + (1 - w_k) w_l \beta_{j,l})} \quad (3)$$

in the following discussion. The details about how to transform other types of inputs are discussed in the literature (Yang, 2001). Subsequently, the activation weights, which reflect the matching degrees of an input vector to all belief rules in the rule base, are calculated by a distributed multiplicative algorithm. Finally, based on the constructed belief decision matrix, the evidential reasoning method is used to infer the estimated output on the given inputs. A brief introduction on the inference steps of BRB systems is provided in Appendix A. Mathematically, according to the analytical ER algorithm in Eq. (A-15), the combined belief degree β_n to the final output is entirely determined by the basic belief degree $\beta_{n,k}$ and the activation weight w_k for each rule.

3. Analysis of evidential reasoning inference mechanism

To facilitate the application of BRB systems, the inference mechanism of the ER approach need to be investigated. Yang and Xu (2002a) studied the typical non-linear reasoning pattern of the ER approach under harmonic, quasi-harmonic, and contradictory evidence. In this section, we provide a more general analysis on the information inference patterns of the ER approach. Two typical belief rules R_k and R_l are used to analyze the inference patterns of a BRB system. If a set of inputs to antecedent attributes activates more than two belief rules, with the recursive nature of the ER approach, we can also assume that R_l is an intermediate belief rule generated by aggregating all other activated belief rules except for R_k . Suppose the belief distributions of the two rules R_k and R_l are given as follows:

$$R_k : \{(D_n, \beta_{n,k}), n = 1, \dots, N\} \text{ with } \sum_{n=1}^N \beta_{n,k} = 1,$$

$$R_l : \{(D_n, \beta_{n,l}), n = 1, \dots, N\} \text{ with } \sum_{n=1}^N \beta_{n,l} = 1.$$

The activation weights are assumed to be w_k and w_l , respectively, and $w_k + w_l = 1$.

According to analytical Eq. (A-15), the combined belief degree β_n ($n = 1, \dots, N$) can be calculated as follows:

This expression exhibits strongly nonlinear relationships among the combined belief degrees and the parameters of the basic belief degrees and the activation weights. In the following we analyze this nonlinear inference mechanism in accordance with the two different types of parameters.

3.1. Monotonic inference to the basic belief degrees

Let us fix all other parameters except $\beta_{n,k}$. Eq. (2) can then be transformed as

It is easily known from Eq. (3) that β_n increases monotonically and β_i ($i \neq n$) decreases monotonically with the increase of $\beta_{n,k}$. In Appendix B.1, we prove that in the analytical ER algorithm the first order derivatives of β_n ($n = 1, \dots, N$) with respect to $\beta_{n,k}$ is not less than zero, i.e., $\frac{\partial \beta_n}{\partial \beta_{n,k}} \geq 0$. We can therefore conclude that β_n increases monotonically with $\beta_{n,k}$.

To illustrate the monotonic reasoning pattern, we consider an illustrative case in which the consequents of the activated rules R_k and R_l are assessed by,

$$R_k : \{(D_{n-1}, \beta_{n-1,k}), (D_n, \beta_{n,k}), (D_{n+1}, \beta_{n+1,k})\} \text{ with } \beta_{n-1,k} + \beta_{n,k} + \beta_{n+1,k} = 1,$$

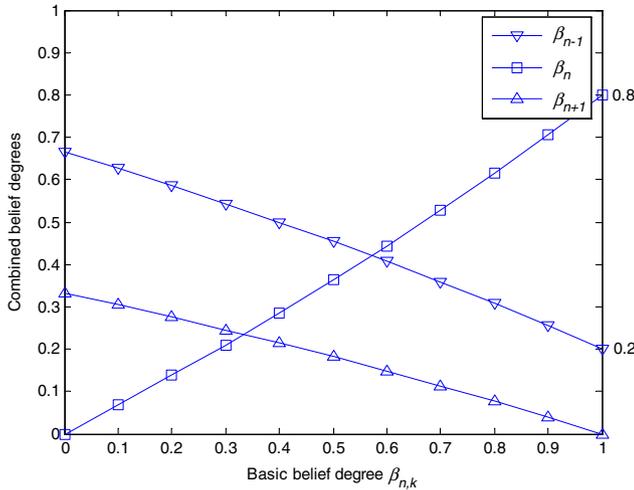


Fig. 2. Illustration of monotonic inference pattern to basic belief degree.

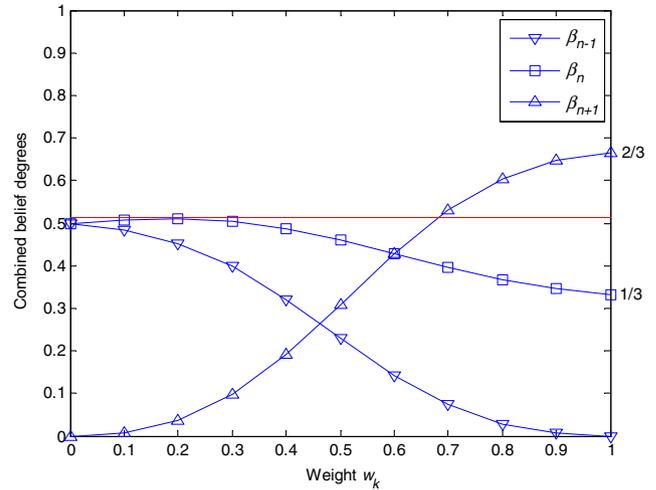


Fig. 3. Illustration of nonlinear approximation process to the activation weight.

$R_l : \{(D_{n-1}, \beta_{n-1,l})\}$ with $D_{n-1,l} = 1$

and $w_k = 2w_l$, $\beta_{n-1,k} = \beta_{n+1,k} = \frac{1}{2}(1 - \beta_{n,k})$. We then have,

$$\beta_{n-1} = \frac{4 - 3\beta_{n,k}}{6 - \beta_{n,k}}, \quad \beta_n = \frac{4\beta_{n,k}}{6 - \beta_{n,k}}, \quad \beta_{n+1} = \frac{2 - 2\beta_{n,k}}{6 - \beta_{n,k}}$$

The curves of β_{n-1} , β_n and β_{n+1} with respect to $\beta_{n,k}$ are shown in Fig. 2.

In Fig. 2, it shows that β_{n-1} and β_{n+1} monotonically decrease and β_n monotonically increases with the increase of $\beta_{n,k}$. The relationships between the combined belief degrees and the basic belief degrees are not a simple linear weighted sum. At $\beta_{n,k} = 0$, the judgements are quasi-harmonic as discussed by Yang and Xu (2002a). We have $\beta_{n-1} = 2/3$, $\beta_n = 0$ and $\beta_{n+1} = 1/3$, which satisfies the synthesis axioms (Yang & Xu, 2002b). At $\beta_{n,k} = 1$, the consequent is completely assessed to H_n , which is contradictory to the consequent in rule R_l , we have $\beta_{n-1} = 0.2$, $\beta_n = 0.8$ and $\beta_{n+1} = 0$, and $\beta_n = (w_k/w_l)^2 \beta_{n-1} = 4\beta_{n-1}$. β_n increases quickly since the activated rule R_k is twice as important as R_l .

3.2. Nonlinear approximation to the activation weights

Suppose $w_k = w$ and note that $w_l = 1 - w$. Eq. (2) can then be transformed as follows:

$$\beta_n = \frac{(\beta_{n,k} + \beta_{n,l} - \beta_{n,k}\beta_{n,l})w^2 + (\beta_{n,k}\beta_{n,l} - 2\beta_{n,l})w + \beta_{n,l}}{(2 - \sum_{j=1}^N \beta_{j,k}\beta_{j,l})w^2 - (2 - \sum_{j=1}^N \beta_{j,k}\beta_{j,l})w + 1} \quad (4)$$

To illustrate the nonlinear approximation process, we consider a specific case in which the consequents of the activation rules R_k and R_l are assessed by,

$$R_k : \{(D_{n-1}, 0), (D_n, 1/3), (D_{n+1}, 2/3)\},$$

$$R_l : \{(D_{n-1}, 1/2), (D_n, 1/2), (D_{n+1}, 0)\}.$$

We then have,

$$\beta_{n-1} = \frac{3w^2 - 6w + 3}{11w^2 - 11w + 6}, \quad \beta_n = \frac{4w^2 - 5w + 3}{11w^2 - 11w + 6},$$

$$\beta_{n+1} = \frac{4w^2}{11w^2 - 11w + 6}$$

Fig. 3 shows the curves of β_{n-1} , β_n and β_{n+1} with respect to weight w . With the increase of w , the combined belief degrees on consequents approximate more closely to the belief degrees of the consequent in rule R_k than that in rule R_l . It is worth noting that

the nonlinear approximation process is not monotonic. In Fig. 3, the combined belief degree β_n initially increases monotonically from $\beta_{n,l} = 1/2$ and then decreases monotonically to $\beta_{n,k} = 1/3$, taking the maximum value of 0.5095 at $w_k = 0.1883$. The nonlinear pattern can also be analyzed using the gradient information, and the first order derivatives with respect to the activation weights can be found in Appendix B.2.

As discussed above, the combined belief degree β_n is determined by the belief degree $\beta_{n,k}$ and the activation weight w_k which is further affected by the rule weight θ_k , attribute weight δ_i and belief degree $\alpha_{i,n}$. For a quantitative attribute x_i , $\alpha_{i,n}$ is transformed from input values using the vector of referential values $A_i = \{A_{i,n}; n = 1, \dots, N_i\}$. So, a change on these parameters including $\beta_{n,k}$, θ_k , δ_i and $A_i = \{A_{i,n}; n = 1, \dots, N_i\}$ can have significant impact on the inference performance of a BRB system. For achieving desirable inference performance, these parameters need to be trained if input–output data are available.

4. Training techniques of BRB systems

For improved inference performance, input–output data sets should be collected to train a BRB system. Fig. 4 shows the functional training framework of BRB systems.

In Fig. 4, \hat{x}_m represents a given input vector; \hat{y}_m is the corresponding observed output of the real system; y_m is the inference output generated by the BRB system; $\xi(P)$ reflects the difference between \hat{y}_m and y_m as defined later, where P is the vector of training parameters, including $\beta_{n,k}$, θ_k , δ_i and A_i . It is desirable that $\xi(P)$ is as small as possible. An optimal training algorithm is designed to adjust the parameters in order to minimize the difference between

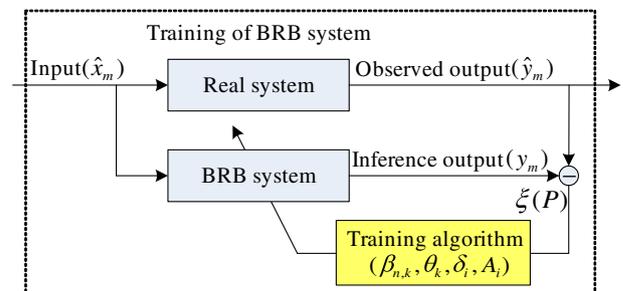


Fig. 4. Training framework of BRB systems.

the observed output and the inference output for the whole set of training data.

4.1. Training data selection

The selection of training data not only affects the inference performance of BRB systems, but also has considerable impact on its applicability. Generally speaking, training data should be selected to cover all possible working conditions of a BRB system. On the other hand, it is also widely believed that comparatively large training data sets may provide better learning performance for intelligent systems (Jang, 1993; Weiss & Provost, 2003). However, the nonlinear inference mechanism of a BRB system makes its training models intractable for large data sets due to the requirements of large storage memory and enormous amounts of training time. Thus, it is crucial to select appropriate data sets for BRB system training.

In a BRB system, belief rules usually decomposes a T -dimensional input space into multiple regional working domains by means of the referential values of antecedent attributes $A_i = \{A_{i,n}; n = 1, \dots, N_i\}$ as shown in Fig. 5. The number of working domains is the Cartesian products of the independent partitions of all antecedent attributes. Take x_i for example. Its partitions consist of the intervals $[A_{i,n_i}, A_{i,n_i+1}]$, $n_i = 1, \dots, N_i - 1$. The working domain starting at the referential point $(A_{1,n_1}, \dots, A_{i,n_i}, \dots, A_{T,n_T})$ can be represented as a T -dimensional subspace $[A_{1,n_1}, A_{1,n_1+1}] \times \dots \times [A_{T,n_T}, A_{T,n_T+1}]$, and 2^T belief rules are associated with this working domain. The total number of working domains is $\prod_{i=1}^T (N_i - 1)$. Each input data in the whole historical data set can be assigned to one or several specific working domains. For example, an input vector transformed from qualitative information may be assigned to several working domains since it is assessed by a distribution. One of the common approaches for training data selection is to use a predetermined percentage to randomly select training data from a data subset associated with a working domain. Further, we can employ the following guidelines or knowledge hidden in belief rule base to adjust the process of sampling training data.

- (1) Select representative training data for all working domains defined by belief rules; make the selected data distribute as evenly as possible in each working domain, and avoid redundant, insignificant and repeated data.
- (2) Increase the amount of training data for more important working domains. Important working domains can be assigned by experts or identified by statistical analysis of historical data.

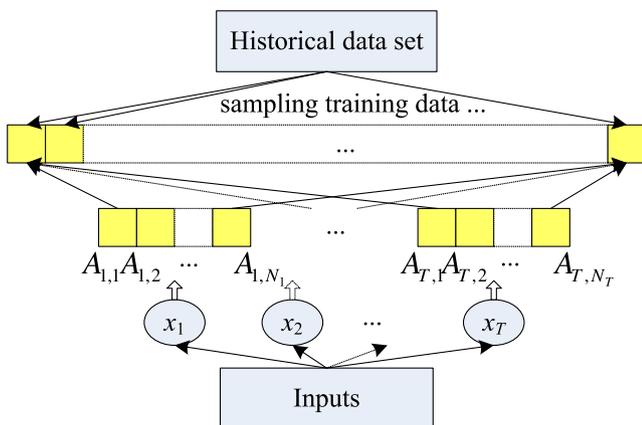


Fig. 5. Training data selection of BRB systems.

- (3) Increase the amount of training data for the working domains in which relevant belief rules have high variation. Let $\beta_n(A_{1,n'_1}, \dots, A_{i,n'_i}, \dots, A_{T,n'_T})$ represent the basic belief degree in the i th belief rule associated with the working domain starting at the referential point $(A_{1,n_1}, \dots, A_{i,n_i}, \dots, A_{T,n_T})$, with n'_i equal to n_i or $n_i + 1$. Without loss of generality, suppose n'_i is equal to n_i and its upward neighboring referential point is $(A_{1,n'_1}, \dots, A_{i,n'_i+1}, \dots, A_{T,n'_T})$. Then the variation of a working domain can be evaluated with the product of “gradient-like” information of its neighboring boundary belief rules as follows,

$$\nabla g(y) = \prod_{v=1}^{2^T} \left| \sum_{n=1}^N u(D_n) \beta_n(A_{1,n'_1}, \dots, A_{i,n'_i+1}, \dots, A_{T,n'_T}) - \sum_{n=1}^N u(D_n) \beta_n(A_{1,n'_1}, \dots, A_{i,n'_i}, \dots, A_{T,n'_T}) \right| \quad (5)$$

4.2. Adaptive training method

As discussed above, belief rules are essentially the sampled input-output points of real systems on the referential values of antecedent attributes. As such, the process of adaptively updating the referential values of antecedent attributes is important for improving the generalization and identification capability of BRB systems. As an extension to the original *local training* model proposed by Yang et al. (2007), in this section an *adaptive training* model is presented to optimize all the parameters in a BRB system including not only belief degrees $\beta_{n,k}$ rule weights θ_k and attribute weights δ_i as considered in the local training model (Yang et al., 2007), but also the referential values of antecedent attributes $A_i = \{A_{i,n}; n = 1, \dots, N_i = |A_i|\}$. According to the physical and functional requirements of BRB systems, the parameters must satisfy the following linear equality and inequality constraints:

- (1) A basic belief degree must not be less than zero or more than one, i.e.,

$$0 \leq \beta_{n,k} \leq 1, \quad n = 1, \dots, N; \quad k = 1, \dots, L \quad (6a)$$

- (2) If the k th belief rule is complete, its total belief degree in the consequents will be equal to one, i.e.,

$$\sum_{n=1}^N \beta_{n,k} = 1, \quad k = 1, \dots, L. \quad (6b)$$

Note that if this equality constraint is required for all the belief rules then the trained belief rule base will be complete.

- (3) A rule weight is normalized, so that it is between zero and one, i.e.,

$$0 \leq \theta_k \leq 1, \quad k = 1, \dots, L. \quad (6c)$$

- (4) An attribute weight is normalized, so that it is between zero and one, i.e.,

$$0 \leq \delta_i \leq 1, \quad i = 1, \dots, T. \quad (6d)$$

- (5) Without loss of generality, suppose x_i is a “profit” attribute, so its referential values must satisfy the preference constraints, i.e.,

$$A_{i,n} - A_{i,n+1} \leq 0, \quad i = 1, \dots, T; \quad n = 1, \dots, N_i - 1 \quad (6e.1)$$

Alternatively, we can also assign a small value V_i to differentiate between two adjacent referential values for attribute x_i . In this case, constraint (6e – 1) can be changed to,

$$A_{i,n} - A_{i,n+1} \leq V_i, \quad i = 1, \dots, T; \quad n = 1, \dots, N_i - 1. \quad (6e.2)$$

In a BRB system, suppose a set of observed training data is provided in the form of M input–output pairs (\hat{x}_m, \hat{y}_m) , $m = 1, \dots, M$, with \hat{x}_m and \hat{y}_m being numerical values. The inference output is represented as belief distribution as described in Eq. (A-13), and its numerical value can be calculated as follows using Eq. (A-14).

$$y_m = \sum_{n=1}^N u(D_n) \beta_n(m),$$

where $\beta_n(m)$ represents the inference output of the BRB system to the m th input. Thus, the objective of the optimal training model is to adjust the basic belief degrees $\beta_{n,k}$ rule weights θ_k , attribute weights δ_i , and referential values of antecedent attributes $A_i = \{A_{i,n}; n = 1, \dots, N_i = |A_i|\}$, which are denoted by $P = P(\beta_{n,k}, \theta_k, \delta_i, A_i)$, in order to minimize the total mean squared error that is defined as follows (Yang et al., 2007),

$$\min_P \xi(P), \quad (7a)$$

where

$$\xi(P) = \frac{1}{M} \sum_{m=1}^M (y_m - \hat{y}_m)^2 \quad (7b)$$

The optimization objective given in Eq. (7b) is obviously a typical nonlinear function. This training model with nonlinear optimization objective and linear equality and inequality constraints can be solved using existing nonlinear optimization software packages, such as the *fmincon* function in the Optimization Toolbox of Matlab (Coleman, Branch, & Grace, 1999).

To illustrate the effects of adaptively training the referential values of antecedent attributes, we use a nonlinear mathematical function to measure the inference performance of the BRB systems without training, with local training and with adaptive training, respectively. The nonlinear function is given as follows:

$$f(x) = x \sin(x^2), \quad 0 \leq x \leq 3 \quad (8)$$

For constructing an illustrative BRB system, we first evenly assign seven referential values within the definition interval of the independent variable, and the vector of referential values is defined by $\{0, 0.5, 1, 1.5, 2, 2.5, 3\}$. For each of the referential points, we can calculate the real output using the above given function, and then transform a numerical output to a belief distribution using the vector of referential values for consequents defined as $\{D_1, D_2, D_3, D_4, D_5\} = \{-2.5, -1, 1, 2, 3\}$. Finally, we can construct the initial belief rule base in which the belief rules are listed in Table 1. Note in Table 1 that the expected value of the consequents of each belief rule is equal to the value of the function at the same input value. For example, in rule no. 1, $y_1 = -2.5 \times 0 + -1 \times 0.5 + 1 \times 0.5 + 2 \times 0 + 3 \times 0 = 0$, and $f(x=0) = 0$, so $y_1 = f(x=0)$.

Subsequently, we generate 1000 training data which are evenly distributed within the definition interval of the input variable. Using the inference methodology discussed in Section 2.3, we can obtain the numerical inference output. Fig. 6 compares the actual output (real line in blue¹) and the inference output (dashed line in red) of the BRB systems without training, with local training and with adaptive training, respectively.

It is evident from Fig. 6 that the inference values using initial belief rule base without training do not match the real values satisfactorily. This means that the initial rule base is not constructed properly although each single belief rule reflects the correct input–output relationship. With local training, the system outputs can minimize the deviation between inference outputs and real out-

Table 1
Initial belief rule base for the nonlinear function.

Rule no.	Rule weight	x	f(x)	Consequents {D ₁ , D ₂ , D ₃ , D ₄ , D ₅ } = {-2.5, -1, 1, 2, 3}
1	1	0	0	{(D ₁ , 0), (D ₂ , 0.5), (D ₃ , 0.5), (D ₄ , 0), (D ₅ , 0)}
2	1	0.5	0.12	{(D ₁ , 0), (D ₂ , 0.44), (D ₃ , 0.56), (D ₄ , 0), (D ₅ , 0)}
3	1	1	0.84	{(D ₁ , 0), (D ₂ , 0.08), (D ₃ , 0.92), (D ₄ , 0), (D ₅ , 0)}
4	1	1.5	1.16	{(D ₁ , 0), (D ₂ , 0), (D ₃ , 0.84), (D ₄ , 0.16), (D ₅ , 0)}
5	1	2	-1.51	{(D ₁ , 0.34), (D ₂ , 0.66), (D ₃ , 0), (D ₄ , 0), (D ₅ , 0)}
6	1	2.5	-0.06	{(D ₁ , 0), (D ₂ , 0.53), (D ₃ , 0.47), (D ₄ , 0), (D ₅ , 0)}
7	1	3	1.2	{(D ₁ , 0), (D ₂ , 0), (D ₃ , 0.8), (D ₄ , 0.2), (D ₅ , 0)}

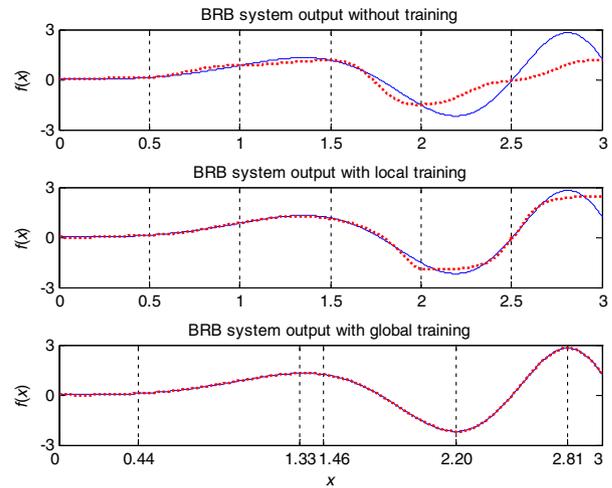


Fig. 6. Comparisons on the BRB systems with different training methods.

puts. However, due to the monotonic inference pattern of evidential reasoning as discussed in Section 3, it is difficult to approximate the real values infinitely in some areas if the referential points are fixed.

However, after the referential values of inputs are adaptively updated using the proposed training method, the inference precision can be improved significantly, and the trained belief rule base can accurately replicate the nonlinear relationship between input and output using the same training data set. Moreover, the BRB system with the adaptive adjustment of referential values is also capable of identifying the critical working points as shown in the bottom subplot of Fig. 6. In the above nonlinear function, the local extreme points, which are regarded as critical points, are 1.355, 2.195, and 2.814 in the defined interval of the input variable. With the adaptive training, all the critical points are closely identified by the updated referential values $\{0, 0.436, 1.333, 1.464, 2.197, 2.807, 3\}$. The trained belief rule base is provided in Table 2.

It should be mentioned in Table 1 that only two adjacent consequents are associated with non-zero belief degrees since the belief distribution of each rule is transformed from a quantitative data, while in Table 2, most consequents of each trained belief rule are with non-zero belief degrees. It means that with training the BRB system can conduct a joint interpolation of all the consequents. This feature of BRB may make it capable of simulating complex systems in a better way than other simple interpolation methods.

5. Numerical studies

In order to illustrate the functionality of BRB systems and validate the effectiveness of the training techniques under study in this paper, in the section we apply BRB to a multi-modal function and a practical pipeline leak detection problem.

¹ For interpretation of color in Fig. 6, the reader is referred to the web version of this article.

Table 2
Trained belief rule base for the nonlinear function.

Rule no.	Rule weight	x	Consequents $\{D_1, D_2, D_3, D_4, D_5\} = \{-2.5, -1, 1, 2, 3\}$
1	1.00	0	$\{(D_1, 0.01), (D_2, 0.53), (D_3, 0.40), (D_4, 0.03), (D_5, 0.03)\}$
2	0.96	0.436	$\{(D_1, 0.06), (D_2, 0.42), (D_3, 0.44), (D_4, 0.02), (D_5, 0.06)\}$
3	0.76	1.333	$\{(D_1, 0.02), (D_2, 0), (D_3, 0.78), (D_4, 0.04), (D_5, 0.16)\}$
4	0.80	1.464	$\{(D_1, 0.15), (D_2, 0.09), (D_3, 0.27), (D_4, 0.05), (D_5, 0.44)\}$
5	0.70	2.197	$\{(D_1, 0.86), (D_2, 0.11), (D_3, 0.03), (D_4, 0), (D_5, 0)\}$
6	0.60	2.807	$\{(D_1, 0), (D_2, 0), (D_3, 0.04), (D_4, 0.10), (D_5, 0.86)\}$
7	0.41	3.00	$\{(D_1, 0.20), (D_2, 0.15), (D_3, 0.03), (D_4, 0), (D_5, 0.62)\}$

5.1. A multi-modal function

Himmelblau function (Himmelblau, 1972) is a multi-modal function commonly used as a benchmark function for testing optimization techniques. In this section we use it to test the inference and identification capability of BRB systems, since lots of inference or prediction problems in engineering systems often involve multiple working modes, which may be regarded as the modals in the multi-modal function. The Himmelblau function is mathematically represented as below (Andrei, 2008),

$$f(x, y) = (x^2 + y - 11)^2 + (x + y^2 - 7)^2, \quad -6 \leq x, y \leq 6. \quad (9)$$

The function has one local maximum $f(x, y) = 181.616$ at point $(-0.270844, -0.923038)$, and four identical local minimums $f(x, y) = 0$ at points $(3.0, 2.0)$, $(3.584428, -1.848126)$, $(-2.805118, 3.131312)$ and $(-3.779310, -3.283186)$. Its contour graph is shown in Fig. 7.

Note that the four local minimums are distributed irregularly on the rugged surface. Without loss of generality, we assign seven referential values $\{-6, -4, -2, 0, 2, 4, 6\}$ for transforming both independent variables, and five referential values $\{0, 200, 500, 1000, 2200\}$ for formulating the consequents. The belief rules can be initially constructed using the real outputs at the 7×7 referential points. Initially, we assume that all the attribute weights and rule weights are equal to 1. To test the inference performance of the BRB system with adaptive training method, we generated 15×5 training data which are evenly distributed on the surface of the multi-modal function. The trained belief rule base for Himmelblau function is given in Appendix table C.1. Fig. 8 shows the real output and the inference output of the trained BRB system.

It is obvious from Fig. 8 that the BRB system updated using the adaptive training method is capable of providing superior data

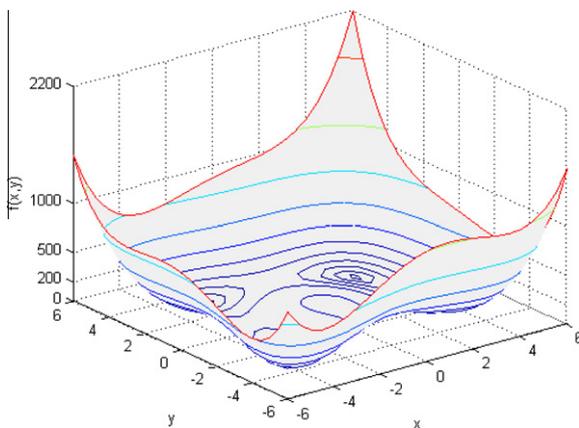


Fig. 7. Contour graph of Himmelblau function.

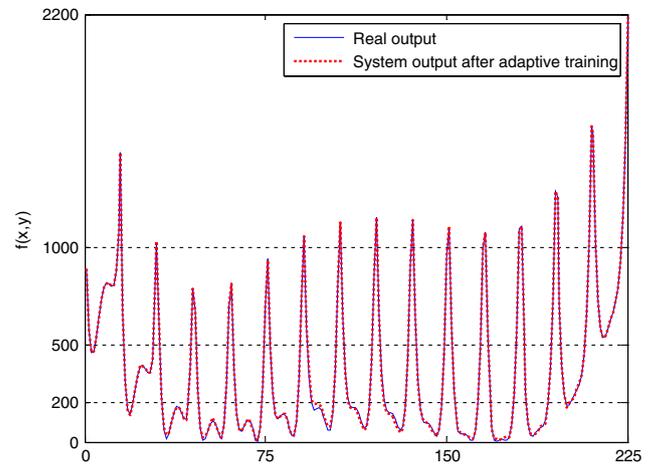


Fig. 8. Inference output of BRB system on Himmelblau function.

fitting performance for this complicated nonlinear multi-modal function. The updated referential values for the two antecedent variables are $\{-6, -3.8813, -2.05381, -1.03, 3.079516, 5.229138, 6\}$ and $\{-6, -4.85368, -3.1657, -0.21, 3.060166, 4.585267, 6\}$, respectively, and the referential points with the format of belief rules approximately identify the critical local maximal and minimal points.

5.2. A practical application: pipeline leak detection

Environmental and personal safety issues and economic losses to pipeline operating companies require the development of effective methods for pipeline leak detection. Based on the basic mass balance principle, a belief rule based expert system was developed for leak detection in a pipeline of more than 100km installed in Britain (Xu et al., 2007).

5.2.1. Problem formulation

In a liquid transmission pipeline, leaks generally cause pressure to change and also create flow discrepancy between inlet and outlet. According to the principle of conservation of mass and historical information, human experts can provide a set of rules to represent typical relationships between the changing patterns of flow and pressure and leak size. For example, if the inlet flow of a pipeline is larger than its outlet flow, and the pressure in the pipeline still decreases although the total content in the pipeline is increasing, then it is highly likely that there is a leak (Xu et al., 2007). In general, leak size is related to the change scale of flow and pressure. Therefore, the difference between inlet flow and outlet flow and the average pipeline pressure change over time, which are denoted as *FlowDiff* and *PressureDiff*, respectively, are selected as the antecedent attributes in inferring the consequent of leak rate, denoted by *LeakSize*. A series of samples with 4%, 16%, and 25% leak rate were generated from a pipeline more than 100 km in length. The values of *FlowDiff* and *PressureDiff* are calculated from the operational data, which are sampled from the mass flow meters at the inlet and outlet and the pressure meters at the inlet, outlet and eight middle points along the pipeline at the rate of 10 s (Xu et al., 2007). The *LeakSize* values are controlled during the leak trial. Fig. 9 shows the *FlowDiff*, *PressureDiff* and *LeakSize* associated with a series of 25% leak trial data collected in about 5 and half hours.

It is obvious that the leak period is clearly marked by the large *FlowDiff*, but the leak size is affected by both *FlowDiff* and *PressureDiff*. In the following study, we will use these data to train and test



Fig. 9. Trial data with controlled LeakSize and associated FlowDiff, PressureDiff.

the BRB system for correctly detecting the leak and precisely estimating the leak sizes.

5.2.2. Constructing an initial belief rule base

- (1) Referential values of the antecedents and consequent
According to the experts' knowledge, the numerical referential values and corresponding linguistic terms for the antecedents and consequent are provided (Xu et al., 2007).

For the antecedent FlowDiff, the following eight referential points are used:

$$H_1 = \{NL, NM, NS, NVS, Z, PS, PM, PL\},$$

where the linguistic terms are negative large (NL), negative medium (NM), negative small (NS), negative very small (NVS), zero (Z), positive small (PS), positive medium (PM), and positive large (PL). The vector of the corresponding numerical referential values is given by,

$$A_1 = \{-10, -5, -3, -1, 0, 1, 2, 3\} \tag{10}$$

For the antecedent PressureDiff, the following seven referential points are used:

$$H_2 = \{NL, NM, NS, Z, PS, PM, PL\},$$

and the vector of the corresponding numerical referential values is given by,

$$A_2 = \{-0.01, -0.005, -0.002, 0, 0.002, 0.005, 0.01\} \tag{11}$$

For the consequent LeakSize, the following five referential points are used:

$$D = \{D_1, D_2, D_3, D_4, D_5\} = \{Z, VS, M, H, VH\},$$

where the linguistic terms are zero (Z), very small (VS), medium (M), high (H) and very high (VH), and the vector of the corresponding discernible referential values is

$$D = \{0, 2, 4, 6, 8\} \tag{12}$$

- (2) Belief rule base
Using the linguistic terms or the equivalent numerical referential values, the belief rule in the BRB system for leak detection can be represented as follows:

IF FlowDiff is $A_1^k \wedge$ PressureDiff is A_2^k ,
THEN LeakSize is $\{(D_1, \beta_{1,k}), (D_2, \beta_{2,k}), \dots, (D_5, \beta_{5,k})\}$,

$$R_k : \left(\sum_{n=1}^5 \beta_{n,k} \leq 1 \right),$$

with rule weight θ_k , and attribute weight $\delta_{1,k}, \delta_{2,k}$,
 $k \in \{1, \dots, L\}$.

Here, A_1^k and A_2^k are referential values as defined in Eqs. (10) and (11). The complete rule base includes 8×7 combinations of the two antecedents. With the practitioner's experiences and historical data, the initial BRB are constructed as shown in Appendix table D.1. To improve the inference performance, further training is necessary to fine tune those parameters in the initial BRB system.

5.2.3. Training and testing of the rule base

In order to select effective training data from the trial data with no leak and 25% leak rate, the statistical distribution on the working domains separated by the initial referential values of antecedents is given in Fig. 10.

According to the training data selection scheme discussed in Section 4.1, 500 training data from a total sample of 2008 are selected for the different working domains. Thus, the input to both the real system and the rule based system is given by,

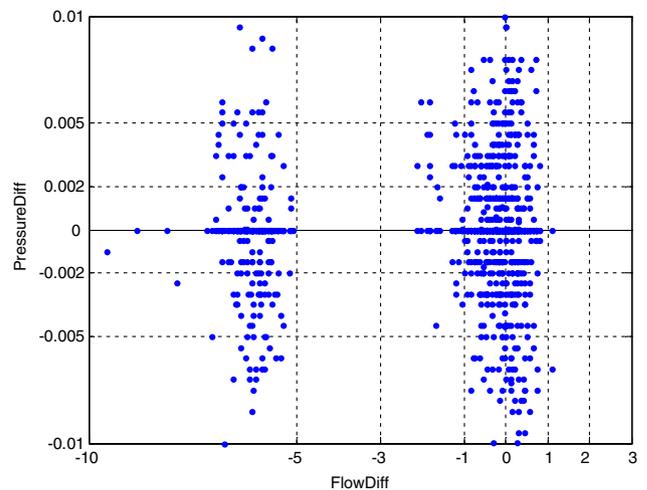


Fig. 10. Statistical distribution of trial data using referential values.

$$X(t) = (FlowDiff(t), PressureDiff(t)).$$

The observed output $\hat{y}(t)$ is the controlled leak size in the trial data. The inference output $y(t)$ is calculated using the inference method discussed in Section 2.3. The training and testing process is outlined as follows.

(1) Set initial parameters of BRB system

The initial degree of beliefs are given by experts and listed in Appendix table D.1. The rule weights $\theta_k (k = 1, \dots, 56)$ and the attribute weights $\delta_i (i = 1, 2)$ are all assumed to be 1, and the referential values of the antecedents are given in Eqs. (10) and (11).

(2) Inference using evidential reasoning

As discussed in Appendix A, each numerical training sample $[FlowDiff(t), PressureDiff(t)]$ can be transformed to belief degrees on the initial referential values using input transformation techniques. The activation weight of each rule is calculated by Eq. (A-3). The activated rules can then be combined using the ER approach. Finally, the estimated output is calculated by Eq. (A-14).

Fig. 11 shows that the values of the estimated *LeakSize* inferred by the initial BRB system are rather different from the sample data. So it is necessary to update the parameters in the initial BRB system.

(3) Train the initial BRB system

As discussed in Section 4.2, the objective of the training process is to find a set of updated parameters P in order to minimize the difference between the observed and estimated $\zeta(P)$, which can be calculated by Eq. (7b). The constraints in the training model are given by Eqs. (6a), (6b), (6c), (6d) and (6e-2) with $V_1 = 1, V_2 = 0.002$. The constrained nonlinear optimization problem can be solved using the nonlinear optimization solver in Matlab. The trained belief rules with rule weights are listed in Appendix table D.2. The updated attribute weights are $\delta_1 = 1$ and $\delta_2 = 0.45$, and the updated referential values are as follows:

$$A_1 = \{-10, -4.1, -2.8, -1.79, -0.79, 0.25, 2, 3\},$$

$$A_2 = \{-0.01, -0.008, -0.005, 0.003, 0.0058, 0.008, 0.01\}.$$

As shown in Fig. 12, the BRB system with adaptive training method can closely replicate the relationship between antecedents *FlowDiff*, *PressureDiff* and consequent *LeakSize* in the selected training data.

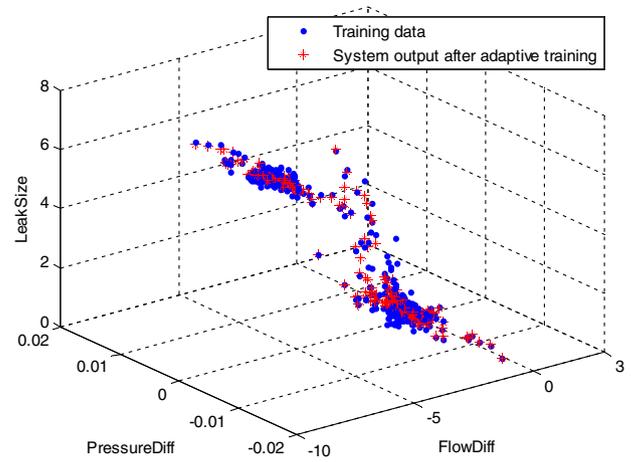


Fig. 12. Training data and the system output after adaptive training.

(4) Test the trained BRB system

To test the trained BRB system, all the 2008 samples in Fig. 9 are used. Fig. 13 shows that the observed *LeakSize* and the estimated *LeakSize* for the same antecedent values. It demonstrates that the estimated outputs match the observed ones very closely.

Fig. 14 displays the observed *LeakSize* and the estimated *LeakSize* on the time scale. It shows that the adaptively trained rule base can clearly detect the leak events and closely predict the leak sizes which happened at around 9:35 a.m and ended at around 10.53 a.m.

It is worth noting that a confirmation period was applied in the trial data for avoiding false leak alarms (Xu et al., 2007). In the testing data in Figs. 13 and 14 samples fall into the confirmation period, which started at around 9:35 a.m and ended at around 9:38 a.m as in Fig. 14.

5.2.4. Comparative analysis and discussions

For the same size of training data, the BRB system updated using the adaptive training method, as studied in this paper, demonstrates much better inference performance than the BRB system updated using the local training method (Xu et al., 2007), and can accurately replicate the complex nonlinear relationship among the antecedents and consequents. In Table 3, the mean absolute error (MAE), root mean squared error (RMSE) and Pearson's correlation coefficient (PCC) between the observed *LeakSize* and the estimated *LeakSize* are used to measure the inference capability. The smaller

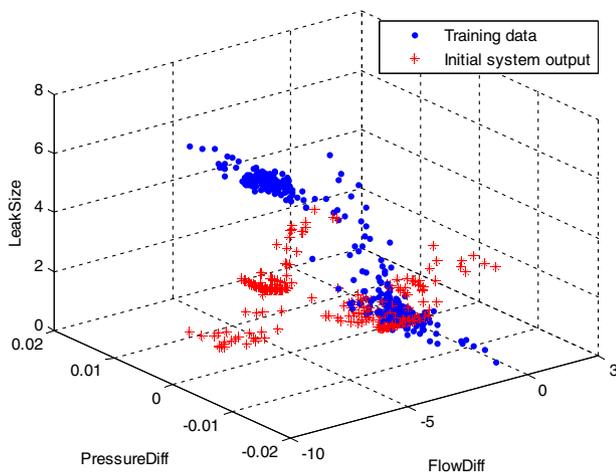


Fig. 11. Training data and the output by the initial BRB system.

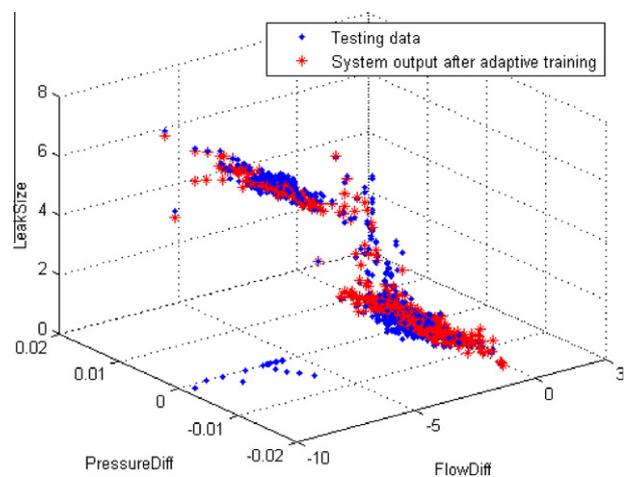


Fig. 13. Testing data and the system output after adaptive training.

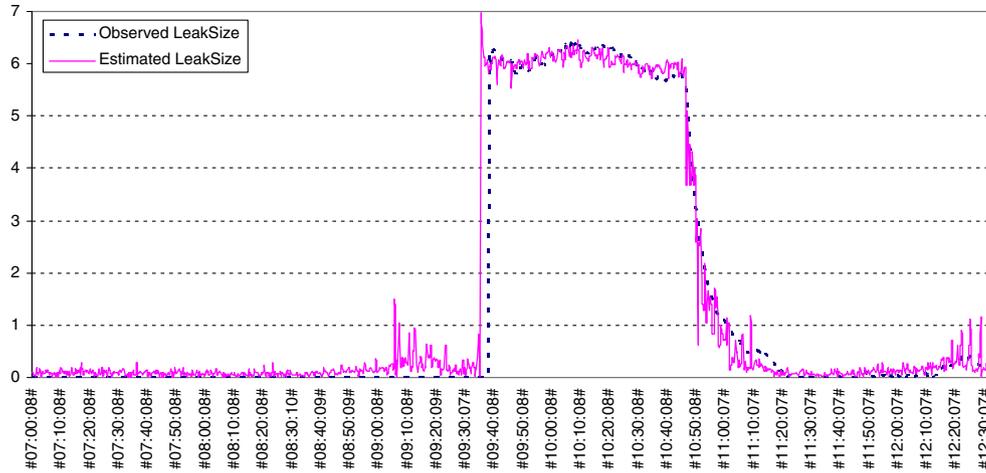


Fig. 14. Testing data and the system output after adaptive training on the time scale.

Table 3

Comparison on local training and adaptive training.

Training algorithms	MAE	RMSE	PCC
Local training	0.22229	0.63791	0.96102
Adaptive training	0.20643	0.63170	0.96697

MAE and RMSE values indicate that the BRB system updated using the adaptive training method has better inference accuracy than the local training method, and the larger PCC value demonstrates that it provides better performance in quickly tracking the changing dynamics of the observed output.

The slight difference between the observed *LeakSize* and the estimated *LeakSize* can be further decreased using noise reduction techniques since the inputs, especially *PressureDiff*, have lots of noise, which is probably caused by dynamic changes in the pipeline and possible data measurement and communication errors (Carpenter, Nicholas, & Henrie, 2005; Silva da, Morooka, Guilherme, Fonseca da, & Mendes, 2005; Xu et al., 2007). The application of noise reduction techniques to improve the performances of BRB systems is not the focus of this paper, and is subject to further research. It should be noted that experts can employ physical principles or prior knowledge to intervene the training process as discussed by Zhou et al. (2009). Also, the computational complexity of the adaptive training method only increases slightly compared with the local training method even if gradient-based optimization techniques are used for training a BRB system, since the referential values just account for a small portion of all the training parameters of the BRB system.

6. Concluding remarks

As an extension of traditional rule based systems, BRB systems provide a more realistic and flexible non-black-box approach for representing human knowledge, aggregating information, and transparently inferring output. The distributed modeling framework employed in a BRB system makes it possible to handle both quantitative data and qualitative information, either complete or incomplete, and enables users to describe antecedent attributes in a flexible, reliable, explicit and systematic manner. The in-depth studies of this paper reveal the inference mechanism of the ER approach and led to the development of the adaptive training method for updating the parameters of BRB systems including the referential values. The main contributions of this paper are summarized as follows:

- (1) The general architecture for single-layer BRB inference systems is organized and its reasoning process is briefly discussed.
- (2) The typical reasoning patterns of the ER approach are examined both experimentally and analytically. These studies reveal the distinctive features of BRB systems, and also demonstrate the necessity for adaptively updating the parameters of BRB systems.
- (3) A heuristic training data selection scheme for BRB systems are proposed. The adaptive training method is proposed to fine tune all parameters of a BRB system, including rule weights, attribute weights, the basic belief degrees of consequents and the referential values of antecedent attributes.

In addition, two numerical studies are conducted to apply the BRB systems to a multi-modal function and a practical pipeline leak detection problem. The superior knowledge representation and inference performances demonstrate that BRB systems have wide potential applications in knowledge-based system development, such as risk analysis, classification, and fault diagnosis

Acknowledgements

This work forms part of the project supported by the UK Engineering and Physical Science Research Council under Grant No. EP/F024606/1 and the Natural Science Foundation of China under Grant No. 60736026.

Appendix A. Introduction to the inference process of BRB systems

The inference of BRB systems mainly includes three steps: (A.1) input transformation, (A.2) calculation of activation weights, and (A.3) inference using evidential reasoning, which is briefly discussed and summarized for the completeness of this paper. More details can be found in the literature (Yang et al., 2006, 2007).

A.1. Input transformation

Suppose $h_{i,n}$ ($n = 1, \dots, N_i$) is the n th possible referential value for attribute x_i . Without loss of generality, let us assume that x_i is a “profit” attribute, which means a large value $h_{i,n+1}$ is preferred to a smaller value $h_{i,n}$. Let h_{i,N_i} be the largest referential value and $h_{i,1}$ the smallest referential value for attribute x_i . Note that these referential values can be assigned directly by experts and further

learned from available input–output data pairs. Given the initial referential values for an attribute, the process of input transformation is thus equivalent to transforming a numerical value into a belief distribution on all the referential values. Specifically, the attribute x_i with an input value h_i can be transformed to the following belief distribution,

$$S(x_i) = \{(h_{i,n}, \alpha_{i,n}), \quad n = 1, \dots, N_i\}.$$

where,

$$\alpha_{i,n} = \frac{h_{i,n+1} - h_i}{h_{i,n+1} - h_{i,n}}, \quad \alpha_{i,n+1} = 1 - \alpha_{i,n} \quad \text{if } h_{i,n} \leq h_i \leq h_{i,n+1},$$

$$\alpha_{i,m} = 0 \quad \text{for } m = 1, \dots, N_i, \text{ and } m \neq n, n + 1.$$

To convert the above piecewise conditional expressions to a unified formula, define two intermediate variables $h_{i,0} = h_{i,N_i}$, $h_{i,N_i+1} = h_{i,1}$, and then the belief degrees of inputs can be calculated as follows:

$$\alpha_{i,n} = \min \left(\max \left(\frac{h_{i,n+1} - h_i}{h_{i,n+1} - h_{i,n}}, 0 \right), \max \left(\frac{h_i - h_{i,n-1}}{h_{i,n} - h_{i,n-1}}, 0 \right) \right), \quad n = 1, \dots, N_i \quad (A.1)$$

Furthermore, suppose the referential value $h_{i,n}$ for attribute x_i is judged to be equivalent to the evaluation grade $H_{i,n}$ in the belief structure, or

$h_{i,n}$ means $H_{i,n}$ ($n = 1, \dots, N_i$).

Then, the multiple inputs can be generally represented as,

$$X = \{(H_{i,n}, \alpha_{i,n}), \quad n = 1, \dots, N_i | i = 1, \dots, T\}, \quad (A.2)$$

where $\alpha_{i,n} \geq 0$ and $\sum_{n=1}^{N_i} \alpha_{i,n} \leq 1$. $\alpha_{i,n}$ represents the likelihood that the input to attribute x_i is assessed to the evaluation grades $H_{i,n}$ or takes the referential value $h_{i,n}$. After all the inputs are transformed into belief distributions, the activation weight on each belief rule can be calculated with a matching algorithm.

A.2. Calculation of activation weights

Generally, in belief rules, the “ \wedge ” connective is used to represent the logical relationship of antecedent attributes. That means the consequent of a rule is not believed to be true unless all the antecedents of the rule are activated. With the belief distributions calculated above, the activation weight w_k of the k th rule is calculated as follows (Yang et al., 2006),

$$w_k = \frac{\theta_k \prod_{i=1}^{T_k} (\alpha_{i,n}^k)^{\bar{\delta}_i}}{\sum_{l=1}^L [\theta_l \prod_{i=1}^{T_l} (\alpha_{i,n}^l)^{\bar{\delta}_i}]} \quad \text{and} \quad \bar{\delta}_i = \frac{\delta_i}{\max_{i=1, \dots, T_k} \{\delta_i\}}, \quad (A.3)$$

where the rule weight θ_k ($k = 1, \dots, L$) and attribute weight δ_i ($i = 1, \dots, T$) are initially provided in the belief rule base. Note that relative weight for each attribute can be different in different belief rules. Because the rule weight θ_k represents the relative importance of the k th rule, without loss of generality, we can also assume that the values of rule weights are normalized, so that

$$0 \leq \theta_k \leq 1 \quad \text{and} \quad \sum_{k=1}^L \theta_k = 1.$$

$\alpha_{i,n}^k$ ($i = 1, \dots, T_k$), which is called the *individual matching degree*, is the belief degree to which the input for the i th antecedent attribute is assessed to the referential value A_i^k in the k th rule. $\theta_k \prod_{i=1}^{T_k} (\alpha_{i,n}^k)^{\bar{\delta}_i}$ is called the *combined matching degree*, which reflects the matching degree between inputs and the referential values for all antecedents in the k th rule. According to the characteristics of power and exponential functions, the combined matching degree is monotonically non-decreasing with the matched belief degree $\alpha_{i,n}^k$, and monotonically

non-increasing with attribute weight $\bar{\delta}_i$ (or δ_i). Note that if $\bar{\delta}_i = 0$, then $(\alpha_{i,n}^k)^{\bar{\delta}_i} = 1$, which means that an antecedent with zero importance does not make any impact on the activation weight; if $\bar{\delta}_i = 1$, then $(\alpha_{i,n}^k)^{\bar{\delta}_i} = \alpha_{i,n}^k$, which means the most important antecedent has significant impact on the activation weight. After normalization, the activation weight w_k is used to measure the degree to which the k th belief rule is weighted and activated. With the calculation of activation weight, the estimated output on a specific input vector can be inferred using the ER approach.

A.3. Inference using evidential reasoning

In a BRB system, suppose that all L rules are independent of each other. The rule base can then be summarized using a belief rule expression matrix (Yang et al., 2007), as show in Table A.1. In the matrix, it is worth noting that the basic belief degree $\beta_{n,k}$ should be updated if incompleteness or ignorance occurs in evaluating the input of antecedents. The details are discussed by Yang et al. (2007). w_k is the activation weight of the k th rule as discussed above.

Based on the belief rule expression matrix, the ER approach can be used to combine activated rules and infer the belief distribution of output. The kernel of the ER approach is a recursive reasoning algorithm which is developed on the basis of and Dempster–Shafer (D–S) theory (Dempster, 1968; Shafer, 1976), fuzzy set theory (Zadeh, 1965), and decision theory (Giovanni & Lurdes, 2009; Yoon & Hwang, 1995). Under the representation scheme of belief structures, the ER approach can simultaneously deal with both quantitative and qualitative information under uncertainty. The reasoning procedure of the ER approach can be summarized as follows.

Firstly, transform the basic belief degree $\beta_{n,k}$ in the belief rule expression matrix into basic probability mass $m_{n,k}$ which represents the degree to which the k th activated rule supports the hypothesis that D_n is the consequent. Let $m_{D,k}$ be the remaining probability mass unassigned to any known consequents. $m_{n,k}$ and $m_{D,k}$ can be calculated from the basic belief degree $\beta_{n,k}$ as follows (Yang & Xu, 2002b),

$$m_{n,k} = w_k \beta_{n,k}, \quad n = 1, \dots, N; \quad k = 1, \dots, L, \quad (A.4)$$

$$m_{D,k} = 1 - \sum_{n=1}^N m_{n,k} = 1 - w_k \sum_{n=1}^N \beta_{n,k}, \quad k = 1, \dots, L. \quad (A.5)$$

Decompose $m_{D,k}$ into $\bar{m}_{D,k}$ and $\tilde{m}_{D,k}$ as follows:

$$\bar{m}_{D,k} = 1 - w_k, \quad \tilde{m}_{D,k} = w_k \left(1 - \sum_{n=1}^N \beta_{n,k} \right), \quad k = 1, \dots, L \quad (A.6)$$

with $m_{D,k} = \bar{m}_{D,k} + \tilde{m}_{D,k}$.

Then, the final output distribution can be inferred using the analytical ER algorithm (Wang, Yang, & Xu, 2006),

Table A-1
Belief rule expression matrix for a BRB system.

Input	Belief output					
	D_1	D_2	\dots	D_n	\dots	D_N
$A_1(w_1)$	$\beta_{1,1}$	$\beta_{2,1}$	\dots	$\beta_{n,1}$	\dots	$\beta_{N,1}$
$A_2(w_2)$	$\beta_{1,2}$	$\beta_{2,2}$	\dots	$\beta_{n,2}$	\dots	$\beta_{N,2}$
\dots						
$A_k(w_k)$	$\beta_{1,k}$	$\beta_{2,k}$	\dots	$\beta_{n,k}$	\dots	$\beta_{N,k}$
\dots						
$A_L(w_L)$	$\beta_{1,L}$	$\beta_{2,L}$	\dots	$\beta_{n,L}$	\dots	$\beta_{N,L}$

$$\{D_n\}: m_n = \mu \left[\prod_{k=1}^L (m_{n,k} + \tilde{m}_{D,k} + \tilde{m}_{D,k}) - \prod_{k=1}^L (\tilde{m}_{D,k} + \tilde{m}_{D,k}) \right],$$

$$n = 1, \dots, N, \quad (\text{A.7})$$

$$\{D\}: \tilde{m}_D = \mu \left[\prod_{k=1}^L (\tilde{m}_{D,k} + \tilde{m}_{D,k}) - \prod_{k=1}^L (\tilde{m}_{D,k}) \right], \quad (\text{A.8})$$

$$\{D\}: \bar{m}_D = \mu \left[\prod_{k=1}^L (\tilde{m}_{D,k}) \right], \quad (\text{A.9})$$

$$\mu = \left[\sum_{i=1}^N \prod_{k=1}^L (m_{i,k} + \tilde{m}_{D,k} + \tilde{m}_{D,k}) - (N-1) \prod_{k=1}^L (\tilde{m}_{D,k} + \tilde{m}_{D,k}) \right]^{-1}, \quad (\text{A.10})$$

$$\{D_n\}: \beta_n = \frac{m_n}{1 - \bar{m}_D}, \quad n = 1, \dots, N, \quad (\text{A.11})$$

$$\{D\}: \beta_D = \frac{\tilde{m}_D}{1 - \bar{m}_D}, \quad n = 1, \dots, N. \quad (\text{A.12})$$

β_n represents the combined belief degree to which the output is assessed to D_n and β_D represents the remaining belief degree unassigned to any known D_n . If $\sum_{n=1}^N \beta_{n,k} = 1$ for all $k = 1, \dots, L$, then $\beta_D = 0$. Eqs. (A-11) and (A-12) are used to provide a normalization process to assign the remaining belief \bar{m}_D back to all the known consequents. It has been proven that $\sum_{n=1}^N \beta_n + \beta_D = 1$ (Yang & Xu, 2002b).

As a consequence, the output of a BRB system can be represented as the following belief structure,

$$S(y) = \{(D_n, \beta_n), \quad n = 1, \dots, N\} \quad (\text{A.13})$$

Further, if we know the utility $u(D_n)$ for each consequent D_n , the numerical output is equal to,

$$y = \sum_{n=1}^N u(D_n) \beta_n. \quad (\text{A.14})$$

After the substitute of intermediate variable, β_n ($n = 1, \dots, N$) can also be represented as follows:

$$\beta_n = \frac{\prod_{k=1}^L (w_k \beta_{n,k} + 1 - w_k \sum_{i=1}^N \beta_{i,k}) - \prod_{k=1}^L (1 - w_k \sum_{i=1}^N \beta_{i,k})}{\sum_{j=1}^N \prod_{k=1}^L (w_k \beta_{j,k} + 1 - w_k \sum_{i=1}^N \beta_{i,k}) - (N-1) \prod_{k=1}^L (1 - w_k \sum_{i=1}^N \beta_{i,k}) - \prod_{k=1}^L (1 - w_k)} \quad (\text{A.15})$$

The logic behind the approach is that, if the k th rule is activated by inputs and its consequent include D_n with $\beta_{n,k} > 0$, then the overall output must be assessed to D_n to a certain degree. The degree is measured by both the degree to which the k th rule is important to the overall output and the degree to which the antecedents of the k th rule are activated by the input vector.

Appendix B. Derivatives of the analytical ER algorithm

According to the analytical algorithm of the ER approach (Wang et al., 2006), the combined belief degrees β_n can be calculated by Eq. (A-15).

To calculate the first order derivatives, we define the following intermediate variables,

$$\mu = \sum_{j=1}^N \prod_{k=1}^L \left(w_k \beta_{j,k} + 1 - w_k \sum_{i=1}^N \beta_{i,k} \right) - (N-1) \\ \times \prod_{k=1}^L \left(1 - w_k \sum_{i=1}^N \beta_{i,k} \right) - \prod_{k=1}^L (1 - w_k),$$

$$X_j = \prod_{k=1}^L (m_{j,k} + \tilde{m}_{H,k} + \tilde{m}_{H,k}) = \prod_{k=1}^L \left(w_k \beta_{j,k} + 1 - w_k \sum_{i=1}^N \beta_{i,k} \right), \\ j = 1, \dots, N,$$

$$Y = \prod_{k=1}^L (\tilde{m}_{H,k} + \tilde{m}_{H,k}) = \prod_{k=1}^L \left(1 - w_k \sum_{i=1}^N \beta_{i,k} \right),$$

$$Z = \prod_{k=1}^L (\tilde{m}_{H,k}) = \prod_{k=1}^L (1 - w_k).$$

Hence, the above analytical equation can be represented as,

$$\beta_n = \frac{X_n - Y}{\sum_{i=1}^N X_i - (N-1)Y - Z}, \quad n = 1, \dots, N$$

The first order derivatives of β_n ($n = 1, \dots, N$) with respect to the parameters p_v ($v = 1, \dots, L + L \times N$) including basic belief degrees $\beta_{n,k}$ and activation weights w_k can be represented as follows:

$$\frac{\partial \beta_n}{\partial p_v} = \frac{1}{\mu^2} \left[\frac{\partial (X_n - Y)}{\partial p_v} (\sum_{j=1}^N X_j - (N-1)Y - Z) - (X_n - Y) \frac{\partial (\sum_{j=1}^N X_j - (N-1)Y - Z)}{\partial p_v} \right] \\ = \frac{1}{\mu^2} \left[\begin{aligned} & \frac{\partial X_n}{\partial p_v} \sum_{j=1}^N X_j - \frac{\partial X_n}{\partial p_v} (N-1)Y - \frac{\partial X_n}{\partial p_v} Z - \frac{\partial Y}{\partial p_v} \sum_{j=1}^N X_j + \frac{\partial Y}{\partial p_v} (N-1)Y + \frac{\partial Y}{\partial p_v} Z \\ & - X_n \sum_{j=1}^N \frac{\partial X_j}{\partial p_v} + X_n (N-1) \frac{\partial Y}{\partial p_v} + X_n \frac{\partial Z}{\partial p_v} + Y \sum_{j=1}^N \frac{\partial X_j}{\partial p_v} - Y (N-1) \frac{\partial Y}{\partial p_v} - Y \frac{\partial Z}{\partial p_v} \end{aligned} \right] \\ = \frac{1}{\mu^2} \left[\begin{aligned} & \frac{\partial X_n}{\partial p_v} \sum_{j=1}^N X_j - \frac{\partial X_n}{\partial p_v} (N-1)Y - \frac{\partial X_n}{\partial p_v} Z - \frac{\partial Y}{\partial p_v} \sum_{j=1}^N X_j + \frac{\partial Y}{\partial p_v} Z \\ & - X_n \sum_{j=1}^N \frac{\partial X_j}{\partial p_v} + X_n (N-1) \frac{\partial Y}{\partial p_v} + X_n \frac{\partial Z}{\partial p_v} + Y \sum_{j=1}^N \frac{\partial X_j}{\partial p_v} - Y \frac{\partial Z}{\partial p_v} \end{aligned} \right].$$

B.1. Derivatives to the basic belief degrees

We first calculate the first order derivatives of the intermediate variables to the basic belief degree $\beta_{n,k}$.

$$\frac{\partial X_j}{\partial \beta_{n,k}} = \begin{cases} 0 \cdot X_n = a_n \cdot X_n, & j = n \\ \frac{-w_k}{w_k \beta_{j,k} + 1 - w_k \sum_{i=1}^N \beta_{i,k}} \cdot X_j = a_j \cdot X_j, & j \neq n, \end{cases}$$

$$\frac{\partial Y}{\partial \beta_{n,k}} = \frac{-w_k}{1 - w_k \sum_{i=1}^N \beta_{i,k}} \cdot Y = b \cdot Y,$$

$$\frac{\partial Z}{\partial \beta_{n,k}} = 0 \cdot Z = c \cdot Z.$$

Using the above equations, we then have the first order derivatives as follows:

$$\begin{aligned} \frac{\partial \beta_n}{\partial \beta_{n,k}} &= \frac{1}{\mu^2} \left[\frac{\partial X_n}{\partial \beta_{n,k}} \sum_{j=1}^N (X_j - Y) + \frac{\partial X_n}{\partial \beta_{n,k}} (Y - Z) + \sum_{j=1}^N \left(Y \frac{\partial X_j}{\partial \beta_{n,k}} - \frac{\partial Y}{\partial \beta_{n,k}} X_j \right) \right. \\ &\quad \left. - X_n \sum_{j=1}^N \left(\frac{\partial X_j}{\partial \beta_{n,k}} - \frac{\partial Y}{\partial \beta_{n,k}} \right) - \frac{\partial Y}{\partial \beta_{n,k}} (X_n - Z) + \frac{\partial Z}{\partial \beta_{n,k}} (X_n - Y) \right] \\ &= \frac{1}{\mu^2} \left[a_n X_n \sum_{j=1}^N (X_j - Y) + a_n X_n (Y - Z) + Y \sum_{j=1}^N (a_j - b) X_j \right. \\ &\quad \left. - X_n \sum_{j=1}^N (a_j X_j - bY) - bY (X_n - Z) + cZ (X_n - Y) \right] \\ &= \frac{1}{\mu^2} \left[Y \sum_{j=1}^N (a_j - b) X_j - X_n \sum_{j=1}^N (a_j X_j - bY) - bY (X_n - Z) \right]. \end{aligned}$$

$$\begin{aligned} a_j - b &= \begin{cases} 0 + \frac{w_k}{1 - w_k \sum_{i=1}^N \beta_{i,k}} \geq 0, & j = n \\ -\frac{w_k}{w_k \beta_{j,k} + 1 - w_k \sum_{i=1}^N \beta_{i,k} + \frac{w_k}{1 - w_k \sum_{i=1}^N \beta_{i,k}}} \geq 0, & j \neq n, \end{cases} \\ a_j X_j - bY &= \begin{cases} \frac{w_k}{1 - w_k \sum_{i=1}^N \beta_{i,k} Y} \geq 0, & j = n \\ -\frac{w_k}{w_k \beta_{j,k} + 1 - w_k \sum_{i=1}^N \beta_{i,k}} X_j + \frac{w_k}{1 - w_k \sum_{i=1}^N \beta_{i,k}} Y \leq 0, & j \neq n, \end{cases} \\ b &= \frac{-w_k}{1 - w_k \sum_{i=1}^N \beta_{i,k}} \leq 0, \end{aligned}$$

According to the mathematical definition of parameters and intermediate variables, it can be shown that,

$$X_j \geq 0, \quad j = 1, \dots, N,$$

$$Y \geq 0,$$

$$X_n - Z \geq 0.$$

We therefore prove that the first order derivatives of β_n ($n = 1, \dots, N$) with respect to $\beta_{n,k}$ is not less than zero, i.e., $\frac{\partial \beta_n}{\partial \beta_{n,k}} \geq 0$. This means that the combined belief degree β_n increases monotonously with the increase of the basic belief degrees $\beta_{n,k}$.

Table C-1
Trained belief rule base for Himmelblau function.

Rule no.	Rule weight	x	y	Consequents $\{D_1, D_2, D_3, D_4, D_5\} = \{0, 200, 500, 1000, 2200\}$
1	0.41	-6	-6	{{(D1, 0.03), (D2, 0.33), (D3, 0.34), (D4, 0), (D5, 0.3)}}
2	0.48	-6	-4.854	{{(D1, 0.72), (D2, 0.06), (D3, 0), (D4, 0), (D5, 0.22)}}
3	0.46	-6	-3.166	{{(D1, 0.21), (D2, 0.28), (D3, 0.15), (D4, 0.36), (D5, 0)}}
4	0.51	-6	-0.21	{{(D1, 0), (D2, 0.48), (D3, 0 (D4, 0.37)), (D5, 0.15)}}
5	0.42	-6	3.060	{{(D1, 0.65), (D2, 0), (D3, 0), (D4, 0), (D5, 0.35)}}
6	0.27	-6	4.585	{{(D1, 0.23), (D2, 0.4), (D3, 0), (D4, 0), (D5, 0.37)}}
7	0.17	-6	6	{{(D1, 0), (D2, 0.13), (D3, 0.27), (D4, 0), (D5, 0.6)}}
8	1.00	-3.881	-6	{{(D1, 0), (D2, 0.47), (D3, 0), (D4, 0.53), (D5, 0)}}
9	1.00	-3.881	-4.854	{{(D1, 0.91), (D2, 0), (D3, 0), (D4, 0.09), (D5, 0)}}
10	1.00	-3.881	-3.166	{{(D1, 1), (D2, 0), (D3, 0), (D4, 0), (D5, 0)}}
11	0.77	-3.881	-0.21	{{(D1, 0.38), (D2, 0.6), (D3, 0), (D4, 0.02), (D5, 0)}}
12	0.53	-3.881	3.060	{{(D1, 0.99), (D2, 0), (D3, 0), (D4, 0), (D5, 0.01)}}
13	0.39	-3.881	4.585	{{(D1, 0.3), (D2, 0.7), (D3, 0), (D4, 0), (D5, 0)}}
14	0.32	-3.881	6	{{(D1, 0), (D2, 0.73), (D3, 0), (D4, 0), (D5, 0.27)}}
15	0.90	-2.054	-6	{{(D1, 0), (D2, 0.33), (D3, 0), (D4, 0.54), (D5, 0.13)}}
16	1.00	-2.054	-4.854	{{(D1, 0.6), (D2, 0.17), (D3, 0), (D4, 0.18), (D5, 0.05)}}
17	0.78	-2.054	-3.166	{{(D1, 0.91), (D2, 0.06), (D3, 0), (D4, 0), (D5, 0.03)}}
18	1.00	-2.054	-0.21	{{(D1, 0.45), (D2, 0.53), (D3, 0.01), (D4, 0), (D5, 0.01)}}
19	0.93	-2.054	3.060	{{(D1, 1), (D2, 0), (D3, 0), (D4, 0), (D5, 0)}}
20	0.36	-2.054	4.585	{{(D1, 0.24), (D2, 0.76), (D3, 0), (D4, 0), (D5, 0)}}
21	0.19	-2.054	6	{{(D1, 0), (D2, 0.69), (D3, 0), (D4, 0.03), (D5, 0.28)}}
22	0.56	-1.03	-6	{{(D1, 0.53), (D2, 0), (D3, 0), (D4, 0), (D5, 0.47)}}
23	0.72	-1.03	-4.854	{{(D1, 0.43), (D2, 0), (D3, 0), (D4, 0.57), (D5, 0)}}
24	1.00	-1.03	-3.166	{{(D1, 0.58), (D2, 0), (D3, 0.42), (D4, 0), (D5, 0)}}
25	1.00	-1.03	-0.21	{{(D1, 0.03), (D2, 0.97), (D3, 0), (D4, 0), (D5, 0)}}
26	1.00	-1.03	3.060	{{(D1, 0.69), (D2, 0.31), (D3, 0), (D4, 0), (D5, 0)}}
27	0.54	-1.03	4.585	{{(D1, 0.01), (D2, 0.99), (D3, 0), (D4, 0), (D5, 0)}}
28	0.50	-1.03	6	{{(D1, 0), (D2, 0.69), (D3, 0), (D4, 0), (D5, 0.31)}}
29	0.72	3.08	-6	{{(D1, 0), (D2, 0.55), (D3, 0), (D4, 0.02), (D5, 0.43)}}
30	0.43	3.08	-4.854	{{(D1, 0), (D2, 0.88), (D3, 0), (D4, 0.12), (D5, 0)}}
31	0.31	3.08	-3.166	{{(D1, 0.93), (D2, 0.07), (D3, 0), (D4, 0), (D5, 0)}}
32	0.58	3.08	-0.21	{{(D1, 0.98), (D2, 0.02), (D3, 0), (D4, 0), (D5, 0)}}
33	1.00	3.08	3.060	{{(D1, 0.93), (D2, 0.07), (D3, 0), (D4, 0), (D5, 0)}}
34	0.79	3.08	4.585	{{(D1, 0), (D2, 0.97), (D3, 0), (D4, 0), (D5, 0.03)}}
35	0.82	3.08	6	{{(D1, 0.03), (D2, 0.55), (D3, 0), (D4, 0), (D5, 0.42)}}
36	0.09	5.229	-6	{{(D1, 0.46), (D2, 0), (D3, 0), (D4, 0), (D5, 0.54)}}
37	0.08	5.229	-4.854	{{(D1, 0.54), (D2, 0.2), (D3, 0), (D4, 0.02), (D5, 0.24)}}
38	0.06	5.229	-3.166	{{(D1, 0.43), (D2, 0.06), (D3, 0.5), (D4, 0), (D5, 0.01)}}
39	0.20	5.229	-0.21	{{(D1, 0.86), (D2, 0), (D3, 0), (D4, 0), (D5, 0.14)}}
40	0.45	5.229	3.060	{{(D1, 0.46), (D2, 0.39), (D3, 0), (D4, 0), (D5, 0.15)}}
41	0.52	5.229	4.585	{{(D1, 0), (D2, 0.68), (D3, 0.03), (D4, 0), (D5, 0.29)}}
42	0.40	5.229	6	{{(D1, 0.27), (D2, 0), (D3, 0), (D4, 0), (D5, 0.73)}}
43	0.46	6	-6	{{(D1, 0.02), (D2, 0.09), (D3, 0.15), (D4, 0.11 0.63)}}
44	0.46	6	-4.854	{{(D1, 0), (D2, 0.02), (D3, 0.17), (D4, 0.79), (D5, 0.02)}}
45	0.98	6	-3.166	{{(D1, 0.21), (D2, 0.26), (D3, 0.07), (D4, 0.46), (D5, 0)}}
46	0.75	6	-0.21	{{(D1, 0.71), (D2, 0), (D3, 0), (D4, 0), (D5, 0.29)}}
47	0.94	6	3.060	{{(D1, 0.59), (D2, 0), (D3, 0.03), (D4, 0), (D5, 0.38)}}
48	0.73	6	4.585	{{(D1, 0.07), (D2, 0.16), (D3, 0.09), (D4, 0.32), (D5, 0.36)}}
49	0.59	6	6	{{(D1, 0), (D2, 0.01), (D3, 0), (D4, 0), (D5, 0.99)}}

B.2. Derivatives to the activation weights

Similar to Section B.1, we can also deduce the first order derivatives of the intermediate variables to the weight w_k of the k th attribute, and then calculate the first order derivatives of combined belief degrees $\beta_n(n = 1, \dots, N)$ with respect to activation weights w_k .

$$\frac{\partial X_j}{\partial w_k} = \frac{(\beta_{j,k} - \sum_{i=1}^N \beta_{i,k})}{(w_k \beta_{j,k} + 1 - w_k \sum_{i=1}^N \beta_{i,k})} X_j = a_j \cdot X_j,$$

$$\frac{\partial Y}{\partial w_k} = \frac{-\sum_{i=1}^N \beta_{i,k}}{(1 - w_k \sum_{i=1}^N \beta_{i,k})} Y = b \cdot Y,$$

$$\frac{\partial Z}{\partial w_k} = \frac{-1}{(1 - w_k)} Z = c \cdot Z.$$

Appendix C. Trained belief rule base for Himmelblau function

See Table C.1.

Appendix D. Initial and trained BRB systems for pipeline leak detection

See Tables D.1 and D.2.

Table D-1
Initial belief rule base with rule weights for pipeline leak detection.

Rule no.	Rule weight	FlowDiff AND PressureDiff	LeakSize distribution $\{D_1, D_2, D_3, D_4, D_5\} = \{0, 2, 4, 6, 8\}$
1	1	NL AND NL	$\{(D_1, 0), (D_2, 0), (D_3, 0), (D_4, 0), (D_5, 1)\}$
2	1	NL AND NM	$\{(D_1, 0), (D_2, 0), (D_3, 0), (D_4, 0.3), (D_5, 0.7)\}$
3	1	NL AND NS	$\{(D_1, 0), (D_2, 0), (D_3, 0.2), (D_4, 0.8), (D_5, 0)\}$
4	1	NL AND Z	$\{(D_1, 0), (D_2, 0), (D_3, 0.8), (D_4, 0.2), (D_5, 0)\}$
5	1	NL AND PS	$\{(D_1, 0.65), (D_2, 0.35), (D_3, 0), (D_4, 0), (D_5, 0)\}$
6	1	NL AND PM	$\{(D_1, 0.85), (D_2, 0.15), (D_3, 0), (D_4, 0), (D_5, 0)\}$
7	1	NL AND PL	$\{(D_1, 0.95), (D_2, 0.05), (D_3, 0), (D_4, 0), (D_5, 0)\}$
8	1	NM AND NL	$\{(D_1, 0), (D_2, 0), (D_3, 0.1), (D_4, 0.9), (D_5, 0)\}$
9	1	NM AND NM	$\{(D_1, 0), (D_2, 0), (D_3, 0.7), (D_4, 0.3), (D_5, 0)\}$
10	1	NM AND NS	$\{(D_1, 0), (D_2, 0.7), (D_3, 0.3), (D_4, 0), (D_5, 0)\}$
11	1	NM AND Z	$\{(D_1, 0), (D_2, 0.9), (D_3, 0.1), (D_4, 0), (D_5, 0)\}$
12	1	NM AND PS	$\{(D_1, 0.8), (D_2, 0.2), (D_3, 0), (D_4, 0), (D_5, 0)\}$
13	1	NM AND PM	$\{(D_1, 0.9), (D_2, 0.1), (D_3, 0), (D_4, 0), (D_5, 0)\}$
14	1	NM AND PL	$\{(D_1, 0.99), (D_2, 0.01), (D_3, 0), (D_4, 0), (D_5, 0)\}$
15	1	NS AND NL	$\{(D_1, 0), (D_2, 0), (D_3, 0.4), (D_4, 0.6), (D_5, 0)\}$
16	1	NS AND NM	$\{(D_1, 0), (D_2, 0), (D_3, 0.8), (D_4, 0.2), (D_5, 0)\}$
17	1	NS AND NS	$\{(D_1, 0), (D_2, 0.3), (D_3, 0.6), (D_4, 0.1), (D_5, 0)\}$
18	1	NS AND Z	$\{(D_1, 0.1), (D_2, 0.7), (D_3, 0.2), (D_4, 0), (D_5, 0)\}$
19	1	NS AND PS	$\{(D_1, 0.7), (D_2, 0.3), (D_3, 0), (D_4, 0), (D_5, 0)\}$
20	1	NS AND PM	$\{(D_1, 0.9), (D_2, 0.1), (D_3, 0), (D_4, 0), (D_5, 0)\}$
21	1	NS AND PL	$\{(D_1, 1), (D_2, 0), (D_3, 0), (D_4, 0), (D_5, 0)\}$
22	1	NVS AND NL	$\{(D_1, 0), (D_2, 0.1), (D_3, 0.4), (D_4, 0.5), (D_5, 0)\}$
23	1	NVS AND NM	$\{(D_1, 0), (D_2, 0.8), (D_3, 0.2), (D_4, 0), (D_5, 0)\}$
24	1	NVS AND NS	$\{(D_1, 0.2), (D_2, 0.7), (D_3, 0.1), (D_4, 0), (D_5, 0)\}$
25	1	NVS AND Z	$\{(D_1, 1), (D_2, 0), (D_3, 0), (D_4, 0), (D_5, 0)\}$
26	1	NVS AND PS	$\{(D_1, 1), (D_2, 0), (D_3, 0), (D_4, 0), (D_5, 0)\}$
27	1	NVS AND PM	$\{(D_1, 1), (D_2, 0), (D_3, 0), (D_4, 0), (D_5, 0)\}$
28	1	NVS AND PL	$\{(D_1, 1), (D_2, 0), (D_3, 0), (D_4, 0), (D_5, 0)\}$
29	1	Z AND NL	$\{(D_1, 0), (D_2, 0.4), (D_3, 0.6), (D_4, 0), (D_5, 0)\}$
30	1	Z AND NM	$\{(D_1, 0.2), (D_2, 0.7), (D_3, 0.1), (D_4, 0), (D_5, 0)\}$
31	1	Z AND NS	$\{(D_1, 0.4), (D_2, 0.6), (D_3, 0), (D_4, 0), (D_5, 0)\}$
32	1	Z AND Z	$\{(D_1, 1), (D_2, 0), (D_3, 0), (D_4, 0), (D_5, 0)\}$
33	1	Z AND PS	$\{(D_1, 1), (D_2, 0), (D_3, 0), (D_4, 0), (D_5, 0)\}$
34	1	Z AND PM	$\{(D_1, 1), (D_2, 0), (D_3, 0), (D_4, 0), (D_5, 0)\}$
35	1	Z AND PL	$\{(D_1, 1), (D_2, 0), (D_3, 0), (D_4, 0), (D_5, 0)\}$
36	1	PS AND NL	$\{(D_1, 0), (D_2, 0.8), (D_3, 0.2), (D_4, 0), (D_5, 0)\}$
37	1	PS AND NM	$\{(D_1, 0.8), (D_2, 0.2), (D_3, 0), (D_4, 0), (D_5, 0)\}$
38	1	PS AND NS	$\{(D_1, 0.95), (D_2, 0.05), (D_3, 0), (D_4, 0), (D_5, 0)\}$
39	1	PS AND Z	$\{(D_1, 1), (D_2, 0), (D_3, 0), (D_4, 0), (D_5, 0)\}$
40	1	PS AND PS	$\{(D_1, 1), (D_2, 0), (D_3, 0), (D_4, 0), (D_5, 0)\}$
41	1	PS AND PM	$\{(D_1, 1), (D_2, 0), (D_3, 0), (D_4, 0), (D_5, 0)\}$
42	1	PS AND PL	$\{(D_1, 1), (D_2, 0), (D_3, 0), (D_4, 0), (D_5, 0)\}$
43	1	PM AND NL	$\{(D_1, 0.1), (D_2, 0.9), (D_3, 0), (D_4, 0), (D_5, 0)\}$
44	1	PM AND NM	$\{(D_1, 0.3), (D_2, 0.7), (D_3, 0), (D_4, 0), (D_5, 0)\}$
45	1	PM AND NS	$\{(D_1, 0.85), (D_2, 0.15), (D_3, 0), (D_4, 0), (D_5, 0)\}$
46	1	PM AND Z	$\{(D_1, 0.98), (D_2, 0.02), (D_3, 0), (D_4, 0), (D_5, 0)\}$
47	1	PM AND PS	$\{(D_1, 1), (D_2, 0), (D_3, 0), (D_4, 0), (D_5, 0)\}$
48	1	PM AND PM	$\{(D_1, 1), (D_2, 0), (D_3, 0), (D_4, 0), (D_5, 0)\}$
49	1	PM AND PL	$\{(D_1, 1), (D_2, 0), (D_3, 0), (D_4, 0), (D_5, 0)\}$
50	1	PL AND NL	$\{(D_1, 0.9), (D_2, 0.1), (D_3, 0), (D_4, 0), (D_5, 0)\}$
51	1	PL AND NM	$\{(D_1, 0.99), (D_2, 0.01), (D_3, 0), (D_4, 0), (D_5, 0)\}$
52	1	PL AND NS	$\{(D_1, 1), (D_2, 0), (D_3, 0), (D_4, 0), (D_5, 0)\}$
53	1	PL AND Z	$\{(D_1, 1), (D_2, 0), (D_3, 0), (D_4, 0), (D_5, 0)\}$
54	1	PL AND PS	$\{(D_1, 1), (D_2, 0), (D_3, 0), (D_4, 0), (D_5, 0)\}$
55	1	PL AND PM	$\{(D_1, 1), (D_2, 0), (D_3, 0), (D_4, 0), (D_5, 0)\}$
56	1	PL AND PL	$\{(D_1, 1), (D_2, 0), (D_3, 0), (D_4, 0), (D_5, 0)\}$

Table D-II
Trained belief rule base with rule weights for pipeline leak detection.

Rule no.	Rule weight	FlowDiff AND PressureDiff	LeakSize distribution $\{D_1, D_2, D_3, D_4, D_5\} = \{0, 2, 4, 6, 8\}$
1	0.95	NL AND NL	$\{(D_1, 0.07), (D_2, 0.01), (D_3, 0), (D_4, 0.01), (D_5, 0.91)\}$
2	0.95	NL AND NM	$\{(D_1, 0.01), (D_2, 0.03), (D_3, 0), (D_4, 0.26), (D_5, 0.70)\}$
3	0.98	NL AND NS	$\{(D_1, 0), (D_2, 0), (D_3, 0.01), (D_4, 0.48), (D_5, 0.51)\}$
4	1.00	NL AND Z	$\{(D_1, 0.06), (D_2, 0.04), (D_3, 0), (D_4, 0.12), (D_5, 0.78)\}$
5	0.60	NL AND PS	$\{(D_1, 0.12), (D_2, 0.10), (D_3, 0.01), (D_4, 0.08), (D_5, 0.69)\}$
6	0.87	NL AND PM	$\{(D_1, 0), (D_2, 0), (D_3, 0), (D_4, 0.22), (D_5, 0.78)\}$
7	0.89	NL AND PL	$\{(D_1, 0), (D_2, 0), (D_3, 0), (D_4, 0.15), (D_5, 0.85)\}$
8	1.00	NM AND NL	$\{(D_1, 0.05), (D_2, 0.02), (D_3, 0.01), (D_4, 0.86), (D_5, 0.06)\}$
9	0.81	NM AND NM	$\{(D_1, 0.03), (D_2, 0.05), (D_3, 0.50), (D_4, 0.28), (D_5, 0.14)\}$
10	0.89	NM AND NS	$\{(D_1, 0.15), (D_2, 0.15), (D_3, 0.22), (D_4, 0.06), (D_5, 0.42)\}$
11	0.79	NM AND Z	$\{(D_1, 0.28), (D_2, 0.12), (D_3, 0), (D_4, 0.05), (D_5, 0.55)\}$
12	0.74	NM AND PS	$\{(D_1, 0.10), (D_2, 0.11), (D_3, 0.01), (D_4, 0.40), (D_5, 0.38)\}$
13	0.88	NM AND PM	$\{(D_1, 0.23), (D_2, 0.13), (D_3, 0), (D_4, 0.14), (D_5, 0.50)\}$
14	0.84	NM AND PL	$\{(D_1, 0.34), (D_2, 0.10), (D_3, 0), (D_4, 0.06), (D_5, 0.50)\}$
15	1.00	NS AND NL	$\{(D_1, 0), (D_2, 0), (D_3, 0.40), (D_4, 0.60), (D_5, 0)\}$
16	0.90	NS AND NM	$\{(D_1, 0.08), (D_2, 0.06), (D_3, 0.75), (D_4, 0.11), (D_5, 0)\}$
17	0.65	NS AND NS	$\{(D_1, 0), (D_2, 0), (D_3, 0.42), (D_4, 0.21), (D_5, 0.37)\}$
18	0.15	NS AND Z	$\{(D_1, 0.14), (D_2, 0.16), (D_3, 0.01), (D_4, 0.23), (D_5, 0.46)\}$
19	1.00	NS AND PS	$\{(D_1, 0.99), (D_2, 0.01), (D_3, 0), (D_4, 0), (D_5, 0)\}$
20	0.86	NS AND PM	$\{(D_1, 0.69), (D_2, 0.01), (D_3, 0), (D_4, 0.03), (D_5, 0.27)\}$
21	1.00	NS AND PL	$\{(D_1, 0.98), (D_2, 0), (D_3, 0), (D_4, 0), (D_5, 0.02)\}$
22	0.84	NVS AND NL	$\{(D_1, 0.19), (D_2, 0.18), (D_3, 0.31), (D_4, 0.32), (D_5, 0)\}$
23	0.59	NVS AND NM	$\{(D_1, 0.80), (D_2, 0.15), (D_3, 0.04), (D_4, 0), (D_5, 0.01)\}$
24	0.54	NVS AND NS	$\{(D_1, 0.49), (D_2, 0.24), (D_3, 0), (D_4, 0.07), (D_5, 0.20)\}$
25	0.22	NVS AND Z	$\{(D_1, 0.14), (D_2, 0.01), (D_3, 0.01), (D_4, 0.22), (D_5, 0.62)\}$
26	0.56	NVS AND PS	$\{(D_1, 0.25), (D_2, 0.23), (D_3, 0), (D_4, 0.04), (D_5, 0.48)\}$
27	1.00	NVS AND PM	$\{(D_1, 0), (D_2, 0.02), (D_3, 0), (D_4, 0.07), (D_5, 0.91)\}$
28	0.96	NVS AND PL	$\{(D_1, 0.53), (D_2, 0.02), (D_3, 0), (D_4, 0.12), (D_5, 0.33)\}$
29	0.97	Z AND NL	$\{(D_1, 0.99), (D_2, 0.01), (D_3, 0), (D_4, 0), (D_5, 0)\}$
30	0.11	Z AND NM	$\{(D_1, 0.61), (D_2, 0.11), (D_3, 0.04), (D_4, 0.02), (D_5, 0.22)\}$
31	0.10	Z AND NS	$\{(D_1, 0.93), (D_2, 0), (D_3, 0.02), (D_4, 0), (D_5, 0.05)\}$
32	0.60	Z AND Z	$\{(D_1, 0.89), (D_2, 0.03), (D_3, 0.03), (D_4, 0), (D_5, 0.05)\}$
33	0.30	Z AND PS	$\{(D_1, 1), (D_2, 0), (D_3, 0), (D_4, 0), (D_5, 0)\}$
34	0.10	Z AND PM	$\{(D_1, 1), (D_2, 0), (D_3, 0), (D_4, 0), (D_5, 0)\}$
35	0.88	Z AND PL	$\{(D_1, 0.98), (D_2, 0.02), (D_3, 0), (D_4, 0), (D_5, 0)\}$
36	0.91	PS AND NL	$\{(D_1, 1), (D_2, 0), (D_3, 0), (D_4, 0), (D_5, 0)\}$
37	1.00	PS AND NM	$\{(D_1, 1), (D_2, 0), (D_3, 0), (D_4, 0), (D_5, 0)\}$
38	0.97	PS AND NS	$\{(D_1, 1), (D_2, 0), (D_3, 0), (D_4, 0), (D_5, 0)\}$
39	0.58	PS AND Z	$\{(D_1, 0.97), (D_2, 0), (D_3, 0), (D_4, 0), (D_5, 0.03)\}$
40	1.00	PS AND PS	$\{(D_1, 0.97), (D_2, 0), (D_3, 0), (D_4, 0.03), (D_5, 0)\}$
41	0.99	PS AND PM	$\{(D_1, 0.97), (D_2, 0.03), (D_3, 0), (D_4, 0), (D_5, 0)\}$
42	1.00	PS AND PL	$\{(D_1, 1), (D_2, 0), (D_3, 0), (D_4, 0), (D_5, 0)\}$
43	0.92	PM AND NL	$\{(D_1, 0.13), (D_2, 0.87), (D_3, 0), (D_4, 0), (D_5, 0)\}$
44	0.93	PM AND NM	$\{(D_1, 0.33), (D_2, 0.67), (D_3, 0), (D_4, 0), (D_5, 0)\}$
45	0.97	PM AND NS	$\{(D_1, 1), (D_2, 0), (D_3, 0), (D_4, 0), (D_5, 0)\}$
46	1.00	PM AND Z	$\{(D_1, 1), (D_2, 0), (D_3, 0), (D_4, 0), (D_5, 0)\}$
47	1.00	PM AND PS	$\{(D_1, 1), (D_2, 0), (D_3, 0), (D_4, 0), (D_5, 0)\}$
48	1.00	PM AND PM	$\{(D_1, 1), (D_2, 0), (D_3, 0), (D_4, 0), (D_5, 0)\}$
49	1.00	PM AND PL	$\{(D_1, 1), (D_2, 0), (D_3, 0), (D_4, 0), (D_5, 0)\}$
50	1.00	PL AND NL	$\{(D_1, 0.90), (D_2, 0.10), (D_3, 0), (D_4, 0), (D_5, 0)\}$
51	1.00	PL AND NM	$\{(D_1, 0.99), (D_2, 0.01), (D_3, 0), (D_4, 0), (D_5, 0)\}$
52	1.00	PL AND NS	$\{(D_1, 1), (D_2, 0), (D_3, 0), (D_4, 0), (D_5, 0)\}$
53	1.00	PL AND Z	$\{(D_1, 1), (D_2, 0), (D_3, 0), (D_4, 0), (D_5, 0)\}$
54	1.00	PL AND PS	$\{(D_1, 1), (D_2, 0), (D_3, 0), (D_4, 0), (D_5, 0)\}$
55	1.00	PL AND PM	$\{(D_1, 1), (D_2, 0), (D_3, 0), (D_4, 0), (D_5, 0)\}$
56	1.00	PL AND PL	$\{(D_1, 1), (D_2, 0), (D_3, 0), (D_4, 0), (D_5, 0)\}$

References

- Andrei, N. (2008). An unconstrained optimization test functions collection. *Advanced Modeling and Optimization*, 10(1), 147–161.
- Azibi, R., & Vanderpooten, D. (2002). Construction of rule-based assignment models. *European Journal of Operational Research*, 138, 274–293.
- Carpenter, P. S., Nicholas, E., & Henrie, M. (2005). Method gives realistic analysis of leak-detection systems. *Oil & Gas Journal*, 103(11), 53–58.
- Coleman, T., Branch, M. A., & Grace, A. (1999). *Optimization toolbox – For use with MATLAB*. Natick, MA: The Mathworks Inc.
- Davis, R. (1986). Knowledge-based systems. *Science*, 231(4741), 957–963.
- Dempster, A. P. (1968). A generalization of Bayesian inference. *Journal of the Royal Statistical Society, Series B*, 30(2), 205–247.
- Giovanni, P., & Lurdes, I. (2009). *Decision theory: Principles and approaches*. John Wiley & Sons.
- Hayes-Roth, F. (1985). Rule-based systems. *Communications of the ACM*, 28(9), 921–932.
- Himmelblau, D. (1972). *Applied nonlinear programming*. New York: McGraw-Hill.
- Jang, J. S. R. (1993). ANFIS: Adaptive-network-based fuzzy inference system. *IEEE Transactions on Systems, Man, and Cybernetics*, 23(3), 665–685.
- Ligeza, A. (2006). *Logical foundations for rule-based systems*. Berlin, Heidelberg: Springer-Verlag.
- Liu, J., Yang, J. B., Ruan, D., Martinez, L., & Wang, J. (2008). Self-tuning of fuzzy belief rule bases for engineering system safety analysis. *Annals of Operations Research*, 163(1), 143–168.
- Negnevitsky, M. (2002). *Artificial intelligence: A guide to intelligent systems*. England: Addison-Wesley.
- Nilsson, N. J. (1982). *Principles of artificial intelligence*. Berlin: Springer.
- Shafer, G. (1976). *Mathematical theory of evidence*. Princeton, NJ: Princeton University Press.
- Silva da, H. V., Morooka, C. K., Guilherme, I. R., Fonseca da, T. C., & Mendes, J. R. P. (2005). Leak detection in petroleum pipelines using a fuzzy system. *Journal of Petroleum Science and Engineering*, 49(3–4), 223–238.
- Sun, R. (1995). Robust reasoning: integrating rule-based and similarity-based reasoning. *Artificial Intelligence*, 75, 241–295.

- Wang, Y. M., Yang, J. B., & Xu, D. L. (2006). Environmental impact assessment using the evidential reasoning approach. *European Journal of Operational Research*, 174(3), 1885–1913.
- Weiss, G. M., & Provost, F. (2003). Learning when training data are costly: The effect of class distribution on tree induction. *Journal of Artificial Intelligence Research*, 19, 315–354.
- Xu, D. L., Liu, J., Yang, J. B., Liu, G. P., Wang, J., Jenkinson, I., et al. (2007). Inference and learning methodology of belief-rule-based expert system for pipeline leak detection. *Expert Systems with Applications*, 32(1), 103–113.
- Yang, J. B. (2001). Rule and utility based evidential reasoning approach for multiple attribute decision analysis under uncertainty. *European Journal of Operational Research*, 131(1), 31–61.
- Yang, J. B., Liu, J., Wang, J., Sii, H. S., & Wang, H. W. (2006). A belief rule-base inference methodology using the evidential reasoning approach – RIMER. *IEEE Transactions on Systems, Man, and Cybernetics – Part A*, 36(2), 266–285.
- Yang, J. B., Liu, J., Xu, D. L., Wang, J., & Wang, H. W. (2007). Optimization models for training belief rule based systems. *IEEE Transactions on Systems, Man, and Cybernetics – Part A*, 37(4), 569–585.
- Yang, J. B., & Singh, M. G. (1994). An evidential reasoning approach for multiple attribute decision making with uncertainty. *IEEE Transactions on Systems, Man, and Cybernetics*, 24(1), 1–18.
- Yang, J. B., & Xu, D. L. (2002a). Nonlinear information aggregation via evidential reasoning in multiattribute decision analysis under uncertainty. *IEEE Transactions on Systems, Man, and Cybernetics – Part A: Systems and Humans*, 32(3), 376–393.
- Yang, J. B., & Xu, D. L. (2002b). On the evidential reasoning algorithm for multiattribute decision analysis under uncertainty. *IEEE Transactions on Systems, Man, and Cybernetics – Part A: Systems and Humans*, 32(3), 289–304.
- Yoon, K., & Hwang, C. L. (1995). *Multiple attribute decision making: An introduction*. Thousand Oaks, CA: Sage Publications.
- Zadeh, L. Z. (1965). Fuzzy sets. *Information Control*, 8(3), 338–353.
- Zhou, Z. J., Hu, C. H., Yang, J. B., Xu, D. L., & Zhou, D. H. (2009). Online updating belief-rule-based system for pipeline leak detection under expert intervention. *Expert Systems with Applications*, 36, 7700–7709.
- Zhou, Z. J., Hu, C. H., Yang, J. B., Xu, D. L., Chen, M. Y., & Zhou, D. H. (2010). A sequential learning algorithm for online constructing belief-rule-based systems. *Expert Systems with Applications*, 37, 1790–1799.