

A sequential learning algorithm for online constructing belief-rule-based systems

Zhi-Jie Zhou^{a,b}, Chang-Hua Hu^a, Jian-Bo Yang^c, Dong-Ling Xu^c, Mao-Yin Chen^b, Dong-Hua Zhou^{b,*}

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

^b Department of Automation, TNLi, Tsinghua University, Beijing 100084, PR China

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

ARTICLE INFO

Keywords:

Belief rule base
Expert system
Evidential reasoning
Recursive algorithm
Utility

ABSTRACT

A belief rule base inference methodology using the evidential reasoning (RIMER) approach has been developed recently. A belief rule base (BRB), which can be treated as a more generalized expert system, extends traditional IF-THEN rules, but requires the assignment of some system parameters including rule weights, attribute weights, and belief degrees. These parameters need to be determined with care for reliable system simulation and prediction. Some off-line optimization models have been proposed, but it is expensive to train and re-train these models in particular for large-scale systems. Moreover, the recursive algorithms are also proposed to fine tune a BRB online, which require less calculation time and satisfy the real-time requirement. However, the earlier mentioned learning algorithms are all based on a predetermined structure of the BRB. For a complex system, prior knowledge may not be perfect, which leads to the construction of an incomplete or even inappropriate initial BRB structure. Also, too many rules in an initial BRB may lead to over fitting, whilst too few rules may result in under fitting. Consequently, such a BRB system may not be capable of achieving overall optimal performance. In this paper, we consider one realistic and important case where both a preliminary BRB structure and system parameters assigned to given rules can be adjusted online. Based on the definition of a new statistical utility for a belief rule as investigated in this paper, a sequential learning algorithm for online constructing more compact BRB systems is proposed. Compared with the other learning algorithms, a belief rule can be automatically added into the BRB or pruned from the BRB, and our algorithm can also satisfy the real-time requirement. In addition, our algorithm inherits the feature of RIMER, i.e., only partial input and output information is required, which could be either incomplete or vague, either numerical or judgmental, or mixed. In order to verify the effectiveness of the proposed algorithm, a practical case study about oil pipeline leak detection is studied and examined to demonstrate how the algorithm can be implemented.

© 2009 Elsevier Ltd. All rights reserved.

1. Introduction

It has become increasingly important to model and analyze decision problems using hybrid information with uncertainty (Walley, 1996; Yang, Liu, Xu, Wang, & Wang, 2007). In order to solve these problems, a generic rule base inference methodology using the evidential reasoning (RIMER) approach was proposed by Yang, Liu, Wang, Sii, and Wang (2006). This methodology establishes a nonlinear relationship between antecedent attributes and an associated consequent, and can reflect the dynamic nature of decision-making problems. The RIMER is developed based on the evidential reasoning (ER) approach (Yang & Singh, 1994; Yang, Wang, Xu, & Chin, 2006), Dempster-Shafer theory of evidence (Dempster, 1968; Shafer, 1976), decision theory (Huang & Yong, 1981), and fuzzy set theory (Zadeh, 1965). It is well known that the IF-THEN rule-based method (Sun, 1995) and the fuzzy IF-THEN

rule-based method (Chen & Saif, 2005; Xu, Liu, Ruan, & Li, 2002) can only cope with fuzzy uncertainty and are not applicable in cases where there exists probabilistic uncertainty. The RIMER approach provides a more informative and flexible scheme than the traditional IF-THEN rule base for knowledge representation, and is capable of capturing vagueness, incompleteness, and nonlinear causal relationships. Equipped with the Windows-based and graphically designed intelligent decision system (IDS) (Yang & Xu, 1999), the RIMER has already been applied to the safety analysis of offshore systems (Liu, Yang, & Sii, 2005).

In the RIMER, the belief rule base (BRB), which can be treated as a more generalized expert system, is used as the tool of knowledge representation. In a BRB, there are several types of system parameters including belief degrees, attribute weights, and rule weights. These parameters need to be determined accurately. However, it is difficult to determine these parameters entirely subjectively, in particular for a large-scale rule base with thousands of rules. Also, a change in a rule weight or an attribute weight may lead to changes in the performance of a BRB (Yang et al., 2007). As such,

* Corresponding author. Tel.: +86 010 62783125; fax: +86 010 62786911.
E-mail address: zdh@mail.tsinghua.edu.cn (D.-H. Zhou).

some optimization models have been proposed to train a BRB (Yang et al., 2007). Because these models are off-line trained and in essence are locally optimal, it is very expensive and time consuming to train and re-train them. In order to solve these problems, the recursive algorithms for online updating the BRB systems have also been developed and they are fast to converge, which is very important for training systems that have a high level of real-time requirement (Zhou, Hu, Yang, Xu, & Zhou, 2009). However, these optimal algorithms are all based on a predetermined structure of BRB (Xu et al., 2007; Zhou et al., 2009).

For a complex system, prior knowledge may not be perfect, which may lead to the construction of an incomplete or even inappropriate initial BRB structure. For example, if there are too many belief rules in an initial BRB, the learning task becomes too complicated to handle, or it is possible to result in over fitting; if there are too few rules in an initial BRB, it may lead to under fitting. Consequently, the system may not be capable of achieving overall optimal performance. To achieve an overall optimal BRB, it is not sufficient to just statistically tune parameters for given rules, but the structure of a BRB system need to be adjusted as well. The final performance of a supervised BRB learning system depends on both its system structure and system parameters (Yang et al., 2007).

Unfortunately, there exists no method to adjust online both the BRB structure and system parameters. Sequential learning algorithms have been investigated in various areas. They can add or statistically prune a neuron for radial basis function (RBF) networks (Huang, Sundararajan, & Saratchandran, 2004, 2005; Lu, Sundararajan, & Saratchandran, 1997, 1998) and fuzzy rules for a fuzzy inference system (FIS) (Rong, Sundararajan, Huang, & Saratchandran, 2006). But these algorithms are not inappropriate in the case where there exists probabilistic and fuzzy uncertainty. Because the RIMER approach can deal with cases where there exists probabilistic and fuzzy uncertainty (Yang et al., 2006, 2006), it is necessary to develop a sequential learning algorithms for online constructing BRB systems.

Based on the definition of the new concept of statistical utility for a belief rule, in this paper, a sequential learning algorithm is proposed for online constructing BRB systems. Compared with existing learning algorithms (Yang et al., 2007; Zhou et al., 2009), a belief rule can be automatically added into a BRB or pruned from the BRB, and our algorithm can also satisfy the real-time requirement. In addition, our algorithm inherits the feature of the RIMER, i.e., only partial input and output information is required, which could be either incomplete or vague, either numerical or judgmental, or mixed. In order to verify the effectiveness of our algorithm, a practical case of oil pipeline leak detection is studied and examined

an investigation into the statistical utility of a belief rule in a BRB. A sequential learning algorithm for online constructing a BRB system is proposed in Section 4. A practical case study of pipeline oil leak detection is presented to verify the algorithms in Section 5. The paper is concluded in Section 6.

2. Preliminaries

2.1. Belief rule base

A belief rule base (BRB), which captures the dynamic of a system, consists of a collection of belief rules defined as follows (Yang et al., 2006):

R_k : If x_1 is $A_1^k \wedge x_2$ is $A_2^k \cdots \wedge x_{M_k}$ is $A_{M_k}^k$,

Then $\{(D_1, \beta_{1,k}), \dots, (D_N, \beta_{N,k})\}$

With a rule weight θ_k and attribute weight $\delta_{1,k}, \delta_{2,k}, \dots, \delta_{M_k,k}$, (1)

where x_1, x_2, \dots, x_{M_k} represents the antecedent attributes in the k th rule. $A_i^k (i = 1, \dots, M_k, k = 1, \dots, L)$ is the referential value of the i th antecedent attribute in the k th rule, $A_i^k \in A_i$, $A_i = \{A_{ij}, j = 1, \dots, J_i\}$ is a set of referential values for the i th antecedent attribute, and J_i is the number of the referential values. $\theta_k (\in \mathbb{R}^+, k = 1, \dots, L)$ is the relative weight of the k th rule, and $\delta_{1,k}, \delta_{2,k}, \dots, \delta_{M_k,k}$ are the relative weights of the M_k antecedent attributes used in the k th rule. $\beta_{j,k} (j = 1, \dots, N, k = 1, \dots, L)$ is the belief degree assessed to D_j , which denotes the j th consequent. If $\sum_{j=1}^N \beta_{j,k} = 1$, the k th rule is said to be complete; otherwise, it is incomplete. Note that “ \wedge ” is a logical connective to represent the “AND” relationship. In addition, suppose that M is the total number of antecedent attributes used in the rule base.

2.2. Belief rule inference using the evidential reasoning approach

When the antecedent attributes, i.e., the inputs of the BRB are available, the evidential reasoning (ER) approach is used as the inference tool. Using the ER analytical algorithms (Wang, Yang, & Xu, 2006), the final conclusion $O(\mathbf{Y}(n))$ that is generated by aggregating all rules, activated by the actual input vector $\hat{\mathbf{x}}(n) = [\hat{x}_1(n), \dots, \hat{x}_M(n)]^T$, can be represented as follows:

$$O(\mathbf{Y}(n)) = h(\hat{\mathbf{x}}(n)) = \{(D_j, \beta_j(\hat{\mathbf{x}}(n))), j = 1, \dots, N\}, \quad (2)$$

where $\beta_j(\hat{\mathbf{x}}(n))$ denotes the belief degree in D_j at time instant n , and Eqs. (3) and (4) hold.

$$\beta_j(\hat{\mathbf{x}}(n)) = \frac{\mu(\hat{\mathbf{x}}(n)) \times \left[\prod_{k=1}^L (\omega_k(\hat{\mathbf{x}}(n)) \beta_{j,k} + 1 - \omega_k(\hat{\mathbf{x}}(n)) \sum_{i=1}^N \beta_{i,k}) - \prod_{k=1}^L (1 - \omega_k(\hat{\mathbf{x}}(n)) \sum_{i=1}^N \beta_{i,k}) \right]}{1 - \mu(\hat{\mathbf{x}}(n)) \times \left[\prod_{k=1}^L (1 - \omega_k(\hat{\mathbf{x}}(n))) \right]}, \quad (3)$$

$$\mu(\hat{\mathbf{x}}(n)) = \left[\sum_{j=1}^N \prod_{k=1}^L \left(\omega_k(\hat{\mathbf{x}}(n)) \beta_{j,k} + 1 - \omega_k(\hat{\mathbf{x}}(n)) \sum_{i=1}^N \beta_{i,k} \right) - (N-1) \prod_{k=1}^L \left(1 - \omega_k(\hat{\mathbf{x}}(n)) \sum_{i=1}^N \beta_{i,k} \right) \right]^{-1}, \quad (4)$$

to demonstrate how our algorithm can be implemented. This shows that our algorithm may be widely applied in engineering.

This paper is organized as follows. In Section 2, some preliminaries about the RIMER approach and rule-based information transformation techniques are briefly reviewed. Section 3 presents

where $\beta_j(\hat{\mathbf{x}}(n))$ is the function of the belief degrees $\beta_{i,k} (i = 1, \dots, N, k = 1, \dots, L)$, the rule weights $\theta_k (k = 1, \dots, L)$, the attribute weights $\delta_i (i = 1, \dots, M)$, and the input vector $\hat{\mathbf{x}}(n)$. $\omega_k(\hat{\mathbf{x}}(n))$, the activation weight of the k th rule at time instant n , can be calculated by

$$\omega_k(\hat{\mathbf{x}}(n)) = \frac{\theta_k \prod_{i=1}^M (\alpha_i^k(\hat{x}_i(n)))^{\bar{\delta}_i}}{\sum_{l=1}^L \theta_l \prod_{i=1}^M (\alpha_i^l(\hat{x}_i(n)))^{\bar{\delta}_i}} \text{ and } \bar{\delta}_i = \frac{\delta_i}{\max_{i=1, \dots, M} \{\delta_i\}}, \quad (5)$$

where $\delta_i \in \mathbb{R}^+$, $i = 1, \dots, M$ is the relative weight of the i th antecedent attribute used in the k th rule. $\alpha_i^k(\hat{x}_i(n)) \in \{\alpha_{ij}(\hat{x}_i(n)), i = 1, \dots, M, j = 1, \dots, J_i\}$, the individual matching degree, is the degree of belief to its j th referential value A_{ij}^k in the k th rule at time instant n . $\alpha_k(\hat{\mathbf{x}}(n)) = \prod_{i=1}^M (\alpha_i^k(\hat{x}_i(n)))^{\bar{\delta}_i}$ is called the normalized combined matching degree.

2.3. Rule-based information transformation technique for quantitative data

In the RIMER, $\alpha_i^k(\hat{x}_i(n))$ could be generated using various ways, depending on the nature of an antecedent attribute and data available such as a qualitative attribute using linguistic values, which is an important characteristic of the RIMER. The input information can be one of the following types: continuous, discrete, symbolic, and ordered symbolic. In order to facilitate data collection, a scheme for handling various types of input information has been summarized by Yang (2001), Yang et al. (2006, 2007). In the proposed scheme, there is an important technique, i.e., rule-based information transformation technique (Yang, 2001), which is used to deal with the input information that includes qualitative assessment and quantitative data. In this paper, we will only consider the quantitative input. So we first review this technique for the quantitative data in this subsection.

Suppose that the input of a quantitative antecedent attribute is given by numerical values. In this case, equivalence rules need to be extracted from the decision maker. This can be used to transform a value to an equivalent expectation, thereby relating a particular value to a set of referential values (Yang et al., 2006). Therefore, a value γ_{ij} ($i = 1, \dots, M; j = 1, \dots, J_i$) can be judged to be a referential value A_{ij} in a BRB, or

$$\gamma_{ij} \text{ means } A_{ij}, \quad i = 1, \dots, M; j = 1, \dots, J_i. \quad (6)$$

Suppose that a large value $\gamma_{i(j+1)}$ is preferred over a small value γ_{ij} . Let γ_{ij_1} and γ_{i1} be the largest and smallest feasible values, respectively. Then, an input value $\hat{x}_i(n)$ is represented using the following equivalent expectation:

$$S(\hat{x}_i(n)) = \{(\gamma_{ij}, \alpha_{ij}(\hat{x}_i(n))), \quad i = 1, \dots, M; j = 1, \dots, J_i\}, \quad (7)$$

where $\alpha_{ij}(\hat{x}_i(n))$ can be calculated by

$$\alpha_{ij}(\hat{x}_i(n)) = \frac{\gamma_{ij+1} - \hat{x}_i(n)}{\gamma_{ij+1} - \gamma_{ij}} \text{ if } \gamma_{ij} \leq \hat{x}_i(n) \leq \gamma_{ij+1}, \quad j = 1, \dots, J_i - 1, \quad (7a)$$

$$\alpha_{ij+1}(\hat{x}_i(n)) = 1 - \alpha_{ij}(\hat{x}_i(n)) \text{ if } \gamma_{ij} \leq \hat{x}_i(n) \leq \gamma_{ij+1}, \quad j = 1, \dots, J_i - 1, \quad (7b)$$

$$\alpha_{is}(\hat{x}_i(n)) = 0 \text{ for } s = 1, \dots, J_i, s \neq j, j + 1. \quad (7c)$$

The quantitative antecedent attribute, $\hat{x}_i(n)$, may also be a random variable and may not always take a single value but several values with different probabilities. In order to solve this problem, the corresponding rule-based information transformation technique has also been proposed by Yang (2001).

3. Statistical utility of a belief rule

In this Section, based on the definition of utility and the neuron's "significance" concept of RBF proposed by Huang et al. (2004, 2005), we will give the statistical utility's definition for a belief rule.

Firstly, according to the definition of the expected utility (Yang, 2001; Yang et al., 2006), the expected utility of $h(\hat{\mathbf{x}}(t))$ on the k th belief rule can be calculated by

$$u_k(\hat{\mathbf{x}}(t)) = \sum_{j=1}^N u(D_j) \omega_k(\hat{\mathbf{x}}(t)) \beta_{j,k}, \quad (8)$$

where $\omega_k(\hat{\mathbf{x}}(t))$ ($k = 1, \dots, L$) is calculated by Eq. (5) and $u(D_j)$ ($j = 1, \dots, N$) denotes the utility of an individual consequence D_j .

Then, according to Eqs. (5) and (8), the average expected utility of $h(\hat{\mathbf{x}}(1)), \dots, h(\hat{\mathbf{x}}(n))$ on the k th belief rule can be given as follows:

$$U(k) = \sum_{j=1}^N u(D_j) \beta_{j,k} \frac{\sum_{t=1}^n c_k(\hat{\mathbf{x}}(t))/n}{\sum_{l=1}^L \sum_{t=1}^n c_l(\hat{\mathbf{x}}(t))/n}, \quad (9)$$

where $c_k(\hat{\mathbf{x}}(t)) = \theta_k \prod_{i=1}^M (\alpha_i^k(\hat{x}_i(t)))^{\bar{\delta}_i}$.

If the limits of the numerator and denominator both exist in Eq. (9), the following equation can be obtained.

$$\lim_{n \rightarrow +\infty} U(k) = \sum_{j=1}^N u(D_j) \beta_{j,k} \frac{\lim_{n \rightarrow +\infty} \sum_{t=1}^n c_k(\hat{\mathbf{x}}(t))/n}{\sum_{l=1}^L \lim_{n \rightarrow +\infty} \sum_{t=1}^n c_l(\hat{\mathbf{x}}(t))/n}. \quad (10)$$

In order to determine $\lim_{n \rightarrow +\infty} U(k)$, we must compute firstly E_k defined by

$$E_k = \lim_{n \rightarrow +\infty} \frac{\sum_{t=1}^n c_k(\hat{\mathbf{x}}(t))}{n}. \quad (11)$$

When the number of the inputs n is large, we obtain the following approximation result.

$$E_k \approx \int_X c_k(\hat{\mathbf{x}}) p(\hat{\mathbf{x}}) d\mathbf{x}, \quad (12)$$

where X denotes the sampling range of $\hat{\mathbf{x}}$ and $p(\hat{\mathbf{x}})$ is the sampling density function.

Furthermore, if $\hat{x}_1(t), \dots, \hat{x}_M(t)$ are assumed to be independent of each other and $p_i(\hat{x}_i)$ is the sampling density function of the i th input in the BRB, we get

$$p(\hat{\mathbf{x}}) = \prod_{i=1}^M p_i(\hat{x}_i). \quad (13)$$

Putting Eq. (13) and $c_k(\hat{\mathbf{x}})$ into Eq. (12) leads to

$$E_k \approx \theta_k \int_X \dots \int_X \prod_{i=1}^M (\alpha_i^k(\hat{x}_i))^{\bar{\delta}_i} \prod_{i=1}^M p_i(\hat{x}_i) d\hat{x}_1 \dots d\hat{x}_M. \quad (14)$$

Suppose that \hat{x}_i ($i = 1, \dots, M$) changes in the interval $[a_i, b_i]$. According to the earlier mentioned independence assumption, there is

$$E_k \approx \theta_k \prod_{i=1}^M I_i^k(\gamma_i^k), \quad (15)$$

where $I_i^k(\gamma_i^k) = \int_{a_i}^{b_i} (\alpha_i^k(\hat{x}_i))^{\bar{\delta}_i} p_i(\hat{x}_i) d\hat{x}_i$, $\gamma_i^k \in \{\gamma_{ij}, i = 1, \dots, M; j = 1, \dots, J_i\}$ is given by Eq. (6) and denotes the referential value of the i th antecedent attribute in the k th belief rule.

From Eq. (14), it is obvious that E_k is dependent on the referential value γ_i^k . So we define $E_k(\gamma_k) \triangleq E_k$, where $\gamma_k = [\gamma_1^k, \dots, \gamma_M^k]^T$ represents the referential vector of the antecedent attributes in the k th belief rule. Thus, when $I_i^k(\gamma_i^k)$ is determined, $E_k(\gamma_k)$ can be obtained by Eq. (15), and the following equation can be obtained once $E_k(\gamma_k)$ is put into Eq. (10).

$$\lim_{n \rightarrow +\infty} U(k) \approx \frac{E_k(0\gamma_k) \sum_{j=1}^N u(D_j) \beta_{j,k}}{\sum_{l=1}^L E_l(\gamma_l)}, \quad (16)$$

where Eq. (16) is considered as the statistical utility of the k th belief rule. If the sampling density function $p(\hat{\mathbf{x}})$ is known, we can determine the statistical utility of a belief rule, which will be investigated under the assumption of the uniform distribution in Appendix A.

The statistical utility of a belief rule gives a measure of the information content in a belief rule and the contribution made by that belief rule to the BRB output over all the input data received so far. It directly links the required learning accuracy to the utility of a belief rule in the learning algorithm so as to realize a compact BRB system. In other words, if the statistical utility of a belief rule is large, it is logical to say that this rule is significant and should be added to a BRB. If an activated belief rule is insignificant in the BRB, it should be pruned. The detail algorithm will be given in Section 4. Therefore, the statistical utility provides a basis to determine the structure of a BRB.

4. A sequential learning algorithm for online constructing a BRB system

In this paper, the proposed sequential learning algorithm for online constructing a BRB system consists of two aspects: structure identification and parameter adjustment. The structure identification mainly includes adding and pruning of a belief rule. The parameter adjustment includes updating the rule weights, the attribute weights, and the belief degrees of a BRB system.

4.1. Adding a belief rule

Firstly, suppose that in an initial BRB, there have been L belief rules, M antecedent attributes, $J_i (i = 1, \dots, M)$ referential values of the i th antecedent attribute, and N consequences, which are kept constant in structure identification. Moreover, suppose that these L belief rules are all significant. Here adding a belief rule can be interpreted as follows: if the criteria as given later are satisfied on the basis of the available input and output information of a BRB system, a new set of referential values of M antecedent attributes will be added and a new belief rule can be constructed.

According to the earlier mentioned statistical utility definition for a belief rule, when a new observed data pair $(\hat{\mathbf{x}}(n), \hat{\mathbf{y}}(n))$ arrives at time instant n , where $\hat{\mathbf{x}}(n)$ is an input vector and $\hat{\mathbf{y}}(n)$ is the corresponding output vector, the following two criteria may be used to determine whether a new belief rule is added.

$$\begin{cases} \|\hat{\mathbf{x}}(n) - \bar{\boldsymbol{\gamma}}(n)\| > \varepsilon, \\ \frac{\hat{\mathbf{y}}(n)E_{L+1}(\gamma_{L+1})}{\sum_{i=1}^{L+1} E_i(\gamma_i)} > e_g, \end{cases} \quad (17)$$

where $\bar{\boldsymbol{\gamma}}(n) = [\bar{\gamma}_1(n), \dots, \bar{\gamma}_M(n)]^T$, $\bar{\gamma}_i(n) \in \{\gamma_{ij}, i = 1, \dots, M; j = 1, \dots, J_i\}$ and $\|\cdot\|$ is the Euclidean norm. $\bar{\boldsymbol{\gamma}}(n)$ denotes the referential vector of the antecedent attributes of a belief rule being nearest to $\hat{\mathbf{x}}(n)$ under the Euclidean distance sense. ε is the distance threshold and e_g is the adding threshold. $L + 1$ is the number of the new belief rule. The output of a BRB, $\hat{\mathbf{y}}(n)$, may be either measured using instruments or assessed by experts, so the output can be either numerical or judgmental. When $\hat{\mathbf{y}}(n)$ is numerical, it can be directly used in Eq. (17); When $\hat{\mathbf{y}}(n)$ is judgmental, it can be represented as $\hat{\mathbf{y}}(n) = \{(D_j, \beta_j(n)), j = 1, \dots, N\}$, and determined by (Yang et al., 2007)

$$\hat{\mathbf{y}}(n) = \sum_{j=1}^N u(D_j) \hat{\beta}_j(n), \quad (17a)$$

where $\hat{\beta}_j(n)$ denotes the degree of belief to which D_j is assessed for the observed data at time instant n .

In Eq. (17), the first criterion requires that a new belief rule may be added if the input data is sufficiently far from the existing belief rule (in the Euclidean distance sense). The second one requires that the statistical utility of a newly added belief rule is greater than a given approximation accuracy.

Once the two criteria given in Eq. (17) are satisfied, a new belief rule, i.e., the $(L + 1)$ th rule, may be added. The parameters of the new rule can be determined as follows:

- (1) The referential value vector of the antecedent attributes is

$$\boldsymbol{\gamma}_{L+1} = \hat{\mathbf{x}}(n). \quad (18a)$$

- (2) The belief degree $\beta_{j,L+1} (j = 1, \dots, N)$, which D_j is assessed for $\hat{\mathbf{y}}(n)$ can be determined using Rule-based information transformation technique (Yang, 2001):

$$\beta_{j,L+1} = \frac{u(D_{j+1}) - \hat{\mathbf{y}}(n)}{u(D_{j+1}) - u(D_j)} \text{ if } u(D_j) \leq \hat{\mathbf{y}}(n) \leq u(D_{j+1}), j = 1, \dots, N - 1, \quad (18b)$$

$$\beta_{j+1,L+1} = 1 - \beta_{j,L+1} \text{ if } u(D_j) \leq \hat{\mathbf{y}}(n) \leq u(D_{j+1}), j = 1, \dots, N - 1, \quad (18c)$$

$$\beta_{s,L+1} = 0 \text{ for } s = 1, \dots, N, s \neq j, j + 1. \quad (18d)$$

- (3) The weights of the antecedent attributes $\bar{\delta}_i (i = 1, \dots, M)$ in the $(L + 1)$ th belief rule are all assigned to be the same as in the other rules. The rule weight can be set as

$$\theta_{L+1} = 1. \quad (18e)$$

The parameters set to the newly added belief rule by Eqs. (18a)–(18e) ensure that with the input $\hat{\mathbf{x}}(n)$, the output of the newly constructed BRB is $\hat{\mathbf{y}}(n)$. It means that the BRB can replicate the relationship between the current input and output more accurately after a new belief rule is added.

4.2. Pruning of a belief rule

If the statistical utility of the k th belief rule is less than the given threshold e_p , i.e., this rule is insignificant, it should be removed. The criterion to prune a belief rule can be described by

$$\frac{E_k(\gamma_k) \sum_{j=1}^N u(D_j) \beta_{j,k}}{\sum_{l=1}^L E_l(\gamma_l)} < e_p, \quad (19)$$

where e_p is the pruning threshold.

In Eqs. (18) and (19), the parameters ε , e_g , and e_p should be chosen appropriately in advance. These parameters can be determined according to the following experience.

- (1) The distance threshold ε is set to around 10% of the upper bound of input variables.
- (2) The adding threshold e_g is chosen according to the desired accuracy. In general, the pruning threshold e_p is set to around 10% of e_g .
- (3) Obviously, if ε and e_g are small, the system performance will be better, but the resulting BRB's structure is more complex, which is a disadvantage when there is high real-time requirement. Therefore, we should choose these parameters carefully according to the expected system performance.

4.3. Parameter adjustment of a BRB

Once the structure of a BRB is determined using the observed data pair $(\hat{\mathbf{x}}(n), \hat{\mathbf{y}}(n))$ on the basis of the initial BRB, some parameters, such as the rule weights, the attribute weights, and the belief degrees, should be updated using $(\hat{\mathbf{x}}(n), \hat{\mathbf{y}}(n))$. We have recently proposed the recursive algorithms for online updating the parameters of a BRB under numerical and judgmental outputs, respec-

tively (Zhou et al., 2009). So we will outline these algorithms in this subsection and then use them directly.

It has been proved that the probabilistic representation of belief is the only appropriate representation of belief, which acts correctly under Dempster's combination rule (Halpern & Fagin, 1992). Since evidence is represented as belief distributions, belief is represented as probability and the Dempster's combination rule is adopted in the ER approach (Yang & Singh, 1994; Yang et al., 2006). It has also been pointed that when the inputs of the BRB are independent, the true outputs, $\hat{\mathbf{y}}(1), \dots, \hat{\mathbf{y}}(n)$, can also be assumed to be independent (Zhou et al., 2009). Therefore, there is

$$f(\hat{\mathbf{y}}(1), \dots, \hat{\mathbf{y}}(n) | \hat{\mathbf{x}}(1), \dots, \hat{\mathbf{x}}(n), \mathbf{Q}) = \prod_{t=1}^n f(\hat{\mathbf{y}}(t) | \hat{\mathbf{x}}(t), \mathbf{Q}), \quad (20)$$

where $\hat{\mathbf{y}}$ is numerical and is considered as a random variable. $f(\hat{\mathbf{y}}(t) | \hat{\mathbf{x}}(t), \mathbf{Q})$ is assumed to be the conditional probability density function (pdf) of $\hat{\mathbf{y}}$ at time instant t and \mathbf{Q} is the unknown parameter vector.

The expectation of the log-likelihood of Eq. (20) at time instant n is defined as

$$L_{n+1}(\mathbf{Q}) \triangleq E \left\{ \sum_{t=1}^n \log f(\hat{\mathbf{y}}(t) | \hat{\mathbf{x}}(t), \mathbf{Q}) \middle| \hat{\mathbf{x}}(1), \dots, \hat{\mathbf{x}}(n), \mathbf{Q}(n) \right\}, \quad (21)$$

where $E\{\cdot\}$ denotes the conditional expectation at $\mathbf{Q} = \mathbf{Q}(n)$.

The recursive formulation of Eq. (21) can be written as

$$L_{n+1}(\mathbf{Q}) = L_n(\mathbf{Q}) + E\{\log f(\hat{\mathbf{y}}(n) | \hat{\mathbf{x}}(n), \mathbf{Q}) | \hat{\mathbf{x}}(n), \mathbf{Q}(n)\}. \quad (22)$$

Define

$$\Gamma_1(\mathbf{Q}(n)) \triangleq \nabla_{\mathbf{Q}} \log f(\hat{\mathbf{y}}(n) | \hat{\mathbf{x}}(n), \mathbf{Q}(n)), \quad (23)$$

$$\Xi_1(\mathbf{Q}(n)) \triangleq E\{-\nabla_{\mathbf{Q}} \nabla_{\mathbf{Q}}^T \log f(\hat{\mathbf{y}}(n) | \hat{\mathbf{x}}(n), \mathbf{Q}) | \hat{\mathbf{x}}(n), \mathbf{Q}(n)\}. \quad (24)$$

Based on the recursive EM algorithm (Chung & Bohme, 2005; Dempster, Laird, & Rubin, 1977; Titterton, 1984), the optimal parameter vector $\mathbf{Q}(n+1)$ is given as follows:

$$\mathbf{Q}(n+1) = \mathbf{Q}(n) + \frac{1}{n} [\Xi_1(\mathbf{Q}(n))]^{-1} \Gamma_1(\mathbf{Q}(n)), \quad (25)$$

where \mathbf{Q} consists of the rule weights, attribute weights and belief degrees satisfying the equality and inequality constraints (Yang et al., 2006, 2007):

$$0 \leq \theta_k \leq 1, \quad k = 1, \dots, L, \quad (25a)$$

$$0 \leq \bar{\delta}_m \leq 1, \quad m = 1, \dots, M, \quad (25b)$$

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

$$\sum_{j=1}^N \beta_{j,k} = 1, \quad k = 1, \dots, L. \quad (25d)$$

Hence, the recursive algorithm (25) can be revised as follows:

$$\mathbf{Q}(n+1) = \prod_{H_1} \left\{ \mathbf{Q}(n) + \frac{1}{n} [\Xi_1(\mathbf{Q}(n))]^{-1} \Gamma_1(\mathbf{Q}(n)) \right\}, \quad (26)$$

where H_1 is a constraint set composed of the constraints (25a)–(25d), and $\prod_{H_1}\{\cdot\}$ is the projection onto the constraint set H_1 , ensuring that the estimation of \mathbf{Q} can satisfy the given constraints. The detail of the algorithm $\prod_{H_1}\{\cdot\}$ has been proposed by Zhou et al. (2009).

If the analytic formulations of $\Xi_1(\mathbf{Q}(n))$ and $\Gamma_1(\mathbf{Q}(n))$ are known, the execution of the algorithm (26) will be less time consuming. So the following assumption is given.

We hope that for a given input, $\hat{\mathbf{x}}(n)$, the BRB system can generate an output, $\mathbf{y}(n)$, as close to $\hat{\mathbf{y}}(n)$ as possible. Here, $\hat{\mathbf{y}}(n)$ is considered as a random variable, and $\mathbf{y}(n)$ can be considered as its

expectation. Hence, we assume that the probability density function (pdf) of $\hat{\mathbf{y}}(n)$ obeys the following normal distribution:

$$f(\hat{\mathbf{y}}(n) | \mathbf{x}(n), \mathbf{Q}) = \frac{1}{\sqrt{2\pi\sigma}} \exp \left\{ -\frac{(\hat{\mathbf{y}}(n) - \mathbf{y}(n))^2}{2\sigma} \right\}, \quad (27)$$

where $\mathbf{Q} = [\mathbf{V}^T, \sigma]^T$ denotes the parameter vector, and σ denotes variance. $\mathbf{V} = [\theta_k, \bar{\delta}_m, \beta_{j,k}]^T$ is parameter vector of the BRB and $k = 1, \dots, L$, $m = 1, \dots, M$ and $j = 1, \dots, N$. The output $\mathbf{y}(n)$ is represented as a distribution, and its average score is given by Yang (2001), Yang et al. (2006, 2007).

$$\mathbf{y}(n) = \sum_{j=1}^N \mu(D_j) \beta_j(\hat{\mathbf{x}}(n)), \quad (28)$$

where $\mu(D_j)$ represents the utility of an individual consequent D_j and $\beta_j(\hat{\mathbf{x}}(n))$ is calculated by Eq. (3).

When Eq. (27) is put into (26), the analytic formulation of the recursive algorithm can be obtained. The detailed algorithm has also been given by Zhou et al. (2009).

When the output of the BRB is judgmental, the similar result is obtained as follows:

$$\mathbf{Q}(n+1) = \prod_{H_2} \left\{ \mathbf{Q}(n) + \frac{1}{n} [\Xi_2(\mathbf{Q}(n))]^{-1} \Gamma_2(\mathbf{Q}(n)) \right\}, \quad (29)$$

where H_2 is a constraint set composed of the constraints (25a)–(25d), and $\prod_{H_2}\{\cdot\}$ is the projection onto the constraint set H_2 . In addition, there are

$$\Gamma_2(\mathbf{Q}(n)) \triangleq \nabla_{\mathbf{Q}} \log f(\hat{\mathbf{B}}(n) | \hat{\mathbf{x}}(n), \mathbf{Q}(n)), \quad (29a)$$

$$\Xi_2(\mathbf{Q}(n)) \triangleq E\{-\nabla_{\mathbf{Q}} \nabla_{\mathbf{Q}}^T \log f(\hat{\mathbf{B}}(n) | \hat{\mathbf{x}}(n), \mathbf{Q}) | \hat{\mathbf{x}}(n), \mathbf{Q}(n)\}, \quad (29b)$$

where $\hat{\mathbf{B}}(n) = [\hat{\beta}_1(n), \dots, \hat{\beta}_N(n)]^T$. $\hat{\beta}_j(n)$ ($j = 1, \dots, N$) is the degree of belief to which D_j is assessed for the observed data at time instant n , and there is $\hat{\mathbf{y}}(n) = \{(D_j, \hat{\beta}_j(n)), j = 1, \dots, N\}$.

Similarly, we also assume that $\hat{\mathbf{B}}$ obeys the complex normal distribution:

$$f(\hat{\mathbf{B}}(n) | \mathbf{x}(n), \mathbf{Q}) = (2\pi)^{-N/2} |\boldsymbol{\chi}|^{-1/2} \times \exp \left\{ -\frac{1}{2} (\hat{\mathbf{B}}(n) - \mathbf{B}(n))^T \boldsymbol{\chi}^{-1} (\hat{\mathbf{B}}(n) - \mathbf{B}(n)) \right\}, \quad (30)$$

where $\mathbf{B}(n) = [\beta_1(n), \dots, \beta_N(n)]^T$ is generated by the BRB system using Eq. (3) for a given input. \mathbf{Q} is the unknown vector and is composed of the parameter vector $\mathbf{V} = [\theta_k, \bar{\delta}_m, \beta_{j,k}]^T$. The entries of the covariance matrix $\boldsymbol{\chi}$ are symmetric positive definite. \mathbf{V} is included in $\mathbf{B}(n)$ and $k = 1, 2, \dots, L$, $m = 1, \dots, M$, $j = 1, \dots, N$.

When Eq. (30) is put into (29), we can also obtain the analytic formulation of the recursive algorithm under the judgmental output.

In the earlier mentioned recursive algorithm, we have the following remarks:

- (1) It is obvious that with some appropriate assumptions, the proposed recursive algorithms are analytic. Also, under some conditions, the algorithms converge to a locally optimal point. Moreover, only part of belief rules are activated at time instant n , so the algorithms can converge fast and can satisfy the real-time requirements.
- (2) Adding a belief rule leads to the increase in the dimensionality of the parameters in a BRB. In other words, the dimensionality of the parameter vector \mathbf{V} will be changed into

$$\mathbf{V} = [\theta_1, \dots, \theta_L, \theta_{L+1}, \bar{\delta}_1, \dots, \bar{\delta}_M, \beta_{1,1}, \dots, \beta_{N,1}, \beta_{1,L}, \dots, \beta_{N,L}, \beta_{1,L+1}, \dots, \beta_{N,L+1}]^T. \quad (31)$$

If the dimensionality of $\mathbf{V}(n)$ changes, $\Gamma_i(\mathbf{Q}(n))(i = 1, 2)$ and $\Xi_i(\mathbf{Q}(n))$ will accordingly change.

4.4. The whole algorithm to construct a BRB system

As a result of the earlier mentioned discussion, the procedure of the proposed learning algorithm for online constructing a BRB system can be summarized as the following steps.

- Step 1: Suppose that in an initial BRB, there have been two belief rules that are both significant. The referential values of the antecedent attributes of the first rule are chosen as $a_i(i = 1, \dots, M)$, and the ones of the second rule are chosen as b_i in the initial BRB, where \hat{x}_i changes in the interval $[a_i, b_i]$. In addition, the appropriate values of ε , e_g and e_p are determined appropriately.
- Step 2: Suppose that there are L belief rules in the BRB, which is updated by the input and output information before time instant n . When the observed data pair, $(\hat{\mathbf{x}}(n), \hat{\mathbf{y}}(n))$, is available at time instant n , the criteria for adding a belief rule in Eq. (17) are checked. If they are satisfied, a new belief rule, i.e., the $(L + 1)$ th rule is added, and the parameters of this rule are determined using Eqs. (18a)–(18e). Otherwise, go to Step 3.
- Step 3: The parameters of the BRB including L belief rules are updated using the recursive algorithms as given in subsection 4.2, and the criteria for pruning a belief rule in Eq. (19) is checked. If the criteria are satisfied, the k th ($k = 1, \dots, L$) rule is removed. Then the dimensionality of the BRB is reduced.
- Step 4: Once the BRB is updated, its knowledge is used to perform inference from the given inputs.

For the earlier mentioned algorithm, the following remarks can be given.

- (1) In Step 1, the significance assumption of the two rules in the initial BRB shows that these rules cannot be pruned, which ensures that $\hat{x}_i(n)$ ($i = 1, \dots, M$) is always located in the range represented by the referential values of the i th antecedent attribute and the rule-based information transformation technique can be used.
- (2) In Step 3, the k th belief rule, satisfying $\omega_k(\hat{\mathbf{x}}(n)) \neq 0$, is called the activated rule. Only the parameters in the activated rule are updated and the statistical utility of this rule changes accordingly. Therefore, only the activated rule should be considered to be checked according to the pruning criteria.
- (3) In the proposed algorithm, based on the belief rules determined by human expert in the initial BRB and the criteria given in Section 4, a belief rule is added or pruned automatically and also the parameters of the BRB system are updated using the training data sequentially (one by one). By this algorithm, there is no need to wait for a long time to collect a complete set of data, which is very important when there is high real-time requirement. Therefore, we can see that this algorithm is sequential and can be used to construct a compact BRB system.

5. A practical case study

In order to present the implementation of the proposed sequential learning algorithm and demonstrate its potential application in engineering, we apply it to build a BRB system for oil pipeline leak detection with data taken from an operational long distance oil pipeline installed in Great Britain (Xu et al., 2007).

5.1. Problem formulation

When a leak develops in a pipeline, some leak data can be obtained. The leak data include the difference between inlet flow and outlet flow, the average pipeline pressure change over time and the leak rate, denoted by *FlowDiff*, *PressureDiff*, and *LeakSize*, respectively. *FlowDiff* and *PressureDiff* are the two very important factors in detecting whether there is leak in the pipeline, and they can be treated as the antecedent attributes of the rule base, and their calculation equations have been proposed by Xu et al. (2007). In addition, *LeakSize* is the consequent attribute of the rule base.

Based on the proposed sequential learning algorithm, we use the data to construct a BRB system for detecting leaks and estimating leak sizes without generating false alarms.

5.2. Online constructing a BRB for leak detection

During the leak trial, 2008 samples of 25% leak data are collected at the rate of 10 s per sample. Figs. 1 and 2 give *FlowDiff* and *PressureDiff*, respectively, when there is no leak and 25% leak. In order to online construct a BRB, 900 data sets are collected in the three periods of 7 a.m.–7:49 a.m., 9:38 a.m.–10:28 a.m., and 10:51 a.m.–11:41 a.m. (Figs. 1 and 2). Then these data are used to construct the BRB using the proposed algorithm. The process of constructing and testing the BRB is implemented as following steps:

Step 1: Set the initial BRB.

According to the prior expert knowledge, it is known that *FlowDiff* and *PressureDiff* changes in the intervals $[-10, 3]$ and $[-0.01, 0.01]$, respectively. In order to construct the initial BRB, we give some linguistic terms as follows: for *FlowDiff*, negative large (NL) and positive large (PL) denote -10 and 3 , respectively; for *PressureDiff*, negative large (NL) and positive large (PL) denote -0.01 and 0.01 , respectively. For the consequent attribute, *LeakSize*, 5 referential points are used: zero (Z), very small (VS), medium (M), high (H), and very high (VH), i.e.,

$$\mathbf{D} = (D_1, D_2, D_3, D_4, D_5) = (Z, VS, M, H, VH). \tag{32}$$

The quantified results of the consequent attribute are given in Table 1. Thus, the initial BRB is constructed as shown in Table 2. We assume that the two belief rules are significant, i.e., they will not be pruned when the BRB is updated. From Figs. 1 and 2, *FlowDiff* and *PressureDiff* are assumed to obey normal and uniform distributions, respectively. In addition, it is assumed that $\varepsilon = 0.6$, $e_g = 0.0005$, and $e_p = 0.00005$.

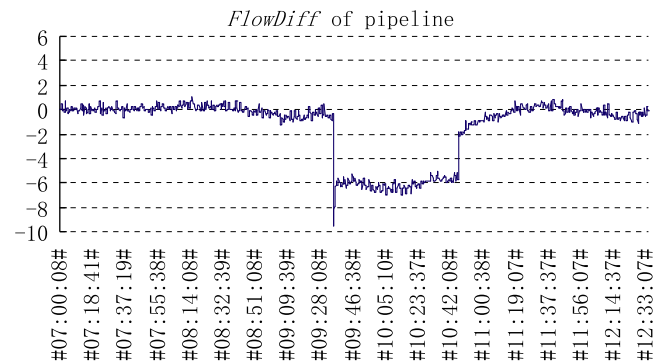


Fig. 1. The *FlowDiff* of the pipeline.

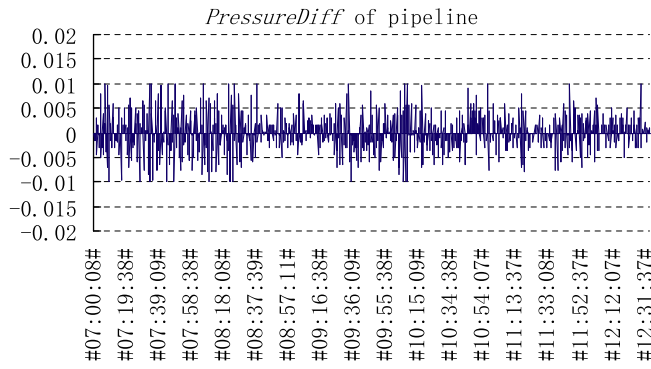


Fig. 2. The PressureDiff of the pipeline.

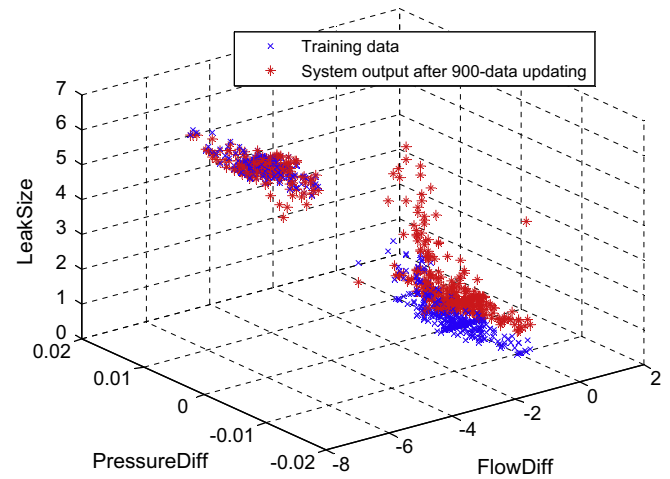


Fig. 4. Training data and the output by the newly constructed BRB.

Table 1
The referential points of LeakSize.

| Linguistic terms | Z | VS | M | H | VH |
|------------------|---|----|---|---|----|
| Numerical values | 0 | 2 | 4 | 6 | 8 |

Table 2
Initial belief rules for pipeline oil leak detection provided by an expert.

| Rule number | FlowDiff AND PressureDiff | LeakSize distribution $\{D_1, D_2, D_3, D_4, D_5\} = \{0, 2, 4, 6, 8\}$ |
|-------------|---------------------------|---|
| 1 | NL AND NL | $\{(D_1, 0), (D_2, 0), (D_3, 0), (D_4, 0), (D_5, 1)\}$ |
| 2 | PL AND PL | $\{(D_1, 1), (D_2, 0), (D_3, 0), (D_4, 0), (D_5, 0)\}$ |

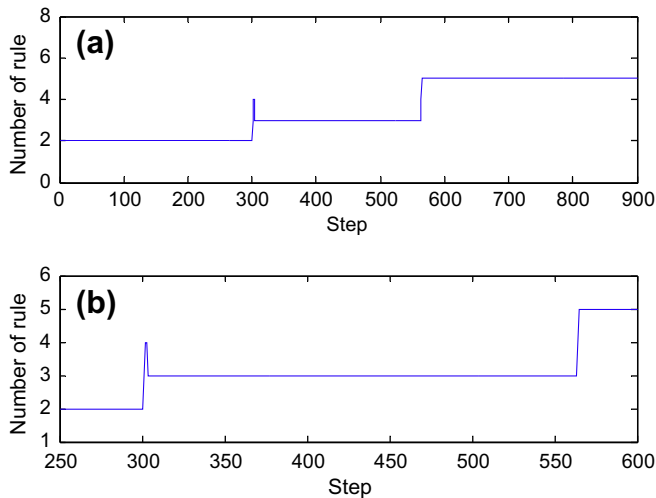


Fig. 3. Number of rules in the BRB.

Step 2: Construct a BRB for leak detection online. Nine hundred training data are used to construct a BRB by the proposed algorithm. Fig. 3a shows the number

of belief rules in the BRB. Fig. 3b gives the more detailed process of online identifying the structure of the BRB from the 250th to 600th training data. From Fig. 3, we can see that the number of the belief rules changes with time and reaches a fixed value at last.

In the process of updating the BRB, except for the referential values of the antecedent attributes that have been provided by an expert in Step 1, some new referential values are added as follows: for FlowDiff, two numerical reference values that are -5.55 and -5.15 are added; for PressureDiff, two numerical reference values that are 0 and 0.001 are added. In order to construct the BRB, equivalence between the numerical value and quality assessment should be provided by the decision maker (Yang, 2001). Thus, -5.55 and -5.15 are qualified as negative medium (NM) and negative small (NS); 0 and 0.001 are qualified as zero (Z) and positive small (PS). Then, the newly constructed BRB which includes five belief rules can be represented in Table 3. In Table 3, the updated rule weight and the belief degrees are also given. Fig. 4 shows that the newly constructed BRB can closely replicate the relationship among FlowDiff, PressureDiff, and LeakSize in the training data. In addition, through calculation, the Mean Squared Error (MSE) (Lewis, 1982) between the trained and the estimated leak data generated by the newly constructed BRB is 0.4199. Compared with the leak data, the MSE is small, which further demonstrate the proposed algorithm.

Step 3: Test the newly constructed BRB.

All the 2008 data shown in Figs. 1 and 2 are used to test the newly constructed BRB as given in Table 3. Fig. 5 gives the observed LeakSize and the estimated LeakSize for the same antecedent values [FlowDiff(t), PressureDiff(t)]. Through calculation, the MSE between the observed and the estimated leak data generated by the

Table 3
Updated belief rules and rule weights of the newly constructed BRB.

| Rule number | Updated rule weight | FlowDiff AND PressureDiff | LeakSize distribution $\{D_1, D_2, D_3, D_4, D_5\} = \{0, 2, 4, 6, 8\}$ |
|-------------|---------------------|---------------------------|---|
| 1 | 0.0009 | NL AND NL | $\{(D_1, 0.2785), (D_2, 0.0372), (D_3, 0), (D_4, 0.0628), (D_5, 6214)\}$ |
| 2 | 1.0000 | NM AND Z | $\{(D_1, 0.0566), (D_2, 0.0288), (D_3, 0.0339), (D_4, 0.7423), (D_5, 0.1384)\}$ |
| 3 | 1.0000 | NS AND Z | $\{(D_1, 0.6768), (D_2, 0.3198), (D_3, 0.0014), (D_4, 0.0010), (D_5, 0.0010)\}$ |
| 4 | 0.6714 | NS AND PS | $\{(D_1, 0.6750), (D_2, 0.2649), (D_3, 0.0003), (D_4, 0.0588), (D_5, 0.0010)\}$ |
| 5 | 1.0000 | PL AND PL | $\{(D_1, 6920), (D_2, 0.2988), (D_3, 0.0082), (D_4, 0.0001), (D_5, 0.0010)\}$ |

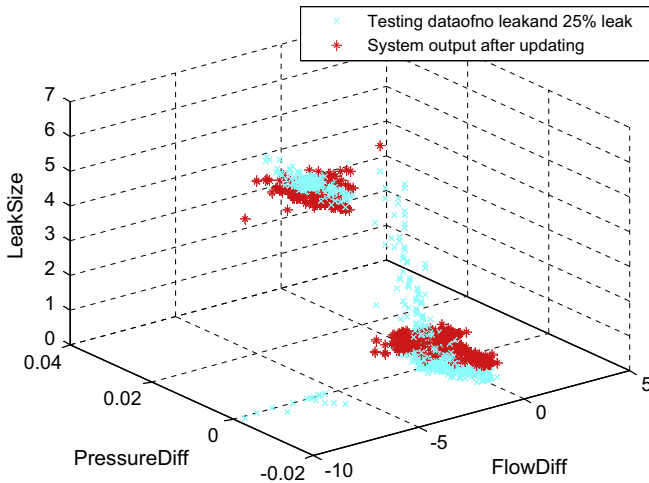


Fig. 5. Testing data of no leak and 25% leak and the output by the newly constructed BRB.

newly constructed BRB is 0.7880. It demonstrates that the estimated outcomes match the observed ones very well. Fig. 6 displays the observed and estimated LeakSize on the time scale. It shows that the newly constructed rule base can detect the leak that happened at around 9:37 a.m. and ended at around 10:49 a.m.

For the earlier mentioned practical case study, we have the following remarks.

- (1) In Table 3, some new referential values of the antecedent attributes are added and all the possible combinations of the referential values, such as the combination “NL AND PL”, are not included. It shows that the BRB is constructed based on the training data, so the belief rules who are inconsistent with the training data will not be added into the BRB. Therefore, only if the training data include all the possible working patterns of the oil pipeline leak, the newly constructed BRB may simulate the real system accurately.
- (2) From Figs. 4 and 5, it can be concluded that the newly constructed BRB can be used to detect oil pipeline leak. In addition, we see that there is noise in the 25% leak detected,

which may be due to noise data recorded from instruments. Therefore, in a real leak detection system, some kind of noise reduction process to smooth data should be included.

5.3. Comparative studies

In order to demonstrate the validity of the proposed algorithm further, the following two aspects in the earlier mentioned case study are compared with the works developed by Xu et al. (2007) and Zhou et al. (2009).

- (1) Because the structure identification is needed in our proposed algorithm, compared with 500 training data used by Xu et al. (2007), 900 training data are used to accomplish the BRB construction for oil leak detection. It shows that the more information is needed in this case study, which is due to the fact that the BRB structure and parameters need to be estimated simultaneously.
- (2) From Fig. 3, we see that the number of belief rules increases or decreases with time, which shows that the structure of the BRB can be identified automatically due to the criteria for adding or pruning a belief rule. Moreover, compared with the BRB, which is composed of 56 belief rules and used for leak detection by Xu et al. (2007) and Zhou et al. (2009), the final BRB, as given in Table 3, constructed by the sequential learning algorithm in this paper only includes five belief rules, which shows that the proposed algorithm can construct the more compact BRB system.

6. Conclusions

This paper is concerned with a sequential learning algorithm for online constructing the belief rule-based systems. The proposed algorithm provides a novel way to estimate the structure and the parameters of a BRB system simultaneously, which is very important for the BRB to achieve the overall optimal performance. Similar to the other optimization models reported for training BRB systems, the proposed algorithm can be used to handle a range of knowledge representation schemes, thereby facilitating the construction of various types of BRB systems, in particular online BRB systems. A practical case study for pipeline oil leak detection is examined to demonstrate how the proposed sequential learning algorithm is implemented, which shows that the proposed algo-

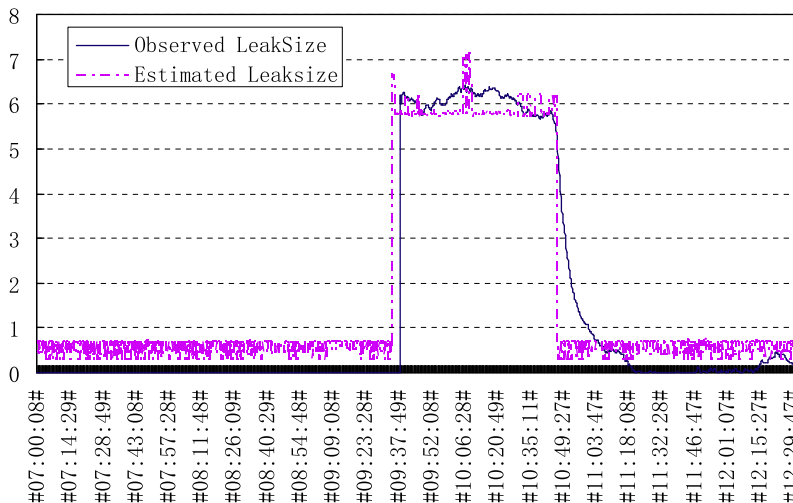


Fig. 6. Testing data of no leak and 25% leak and the output by the newly constructed BRB.

rithm for online constructing the BRB systems may be widely applied in engineering.

There are several features in the proposed algorithm. First of all, using the idea of statistical utility of a belief rule, which is quantitatively defined from a statistical viewpoint as the average information of a belief rule and also the contribution of that belief rule to the overall performance of the BRB system, the proposed algorithm can automatically add or prune a belief rule. This is very helpful to produce a more compact BRB and reduce the system complexity. Secondly, due to an analytical description of relationship between system inputs and outputs that could be discrete or continuous, complete or incomplete, linear, nonlinear or non-smooth, or their mixture (Yang et al., 2006), the proposed recursive algorithm is also analytical, which is very useful to reduce calculation required and satisfies real-time requirements (Zhou et al., 2009). Thirdly, in the proposed algorithm, a belief rule is added or pruned automatically and also the parameters of the BRB system are updated using the training data sequentially (one by one). When new information becomes available, the algorithm immediately adjusts the structure and the parameters of the BRB without having to wait for all information to be provided. Therefore, the algorithm is sequential, which is of great practical significance. Finally but by no means least importantly, the proposed algorithm can be used to process incomplete or vague information, which inherits from the similar feature of RIMER. Equipped with the earlier mentioned features, the sequential learning algorithm is capable of online constructing more compact BRB systems and simulating a range of real systems, especially when there is a high level of real-time requirement and uncertainties.

If the available training data is incomplete (for example, there are missing data), a possible drawback of the proposed algorithm is that the newly constructed BRB might not provide a representative set of rules for simulating the original system. Therefore, only when the training data include all the possible working patterns of the systems, the newly constructed BRB may simulate the originally real system accurately. On the other hand, some thresholds need to be chosen by human experts in the algorithm. The inappropriate thresholds may decrease the accuracy of the newly constructed BRB to simulate the original system.

Although the recursive algorithms for updating the parameters of the BRB systems can converge under some conditions (Zhou et al., 2009) due to the introduction of structure identification, which is an estimating problem of discrete parameters in essence, the convergence of the proposed sequential learning algorithm needs further research.

Acknowledgements

Yang J.B. and Xu D.L. thank the partial support by the UK Engineering and Physical Science Research Council under Grant No.: EP/F024606/1 and by the NSFC under Grant 60736026; Zhou D.H. thanks the partial support by the NSFC under Grant 60721003 and 60736026, and national 973 project under Grant 2009CB32602 and 2010CB731800. We also wish to thank Xu C. for his help in the practical case study.

Appendix A

In order to calculate $I_i^k(\gamma_i^k)$ in Eq. (15), the sampling probability density function $p(\hat{x})$ should be known. In this Appendix, we assume that \hat{x} obeys the uniform distribution.

Under the previously mentioned assumption, $p_i(\hat{x}_i)(i = 1, \dots, M)$ has the following formulation.

$$p_i(\hat{x}_i) = \frac{\hat{x}_i - a_i}{b_i - a_i}, \tag{A.1}$$

where \hat{x}_i changes in the interval $[a_i, b_i]$.

According to Eq. (15), $I_i^k(\gamma_i^k)$ can be determined by

- (1) If $j = 1$, i.e., $\gamma_i^k = \gamma_{ij}$, from Eq. (7a), there is

$$\begin{aligned} I_i^k(\gamma_i^k) &= \int_{\gamma_{ij}}^{\gamma_{ij+1}} \alpha_{ij}(\hat{x}_i)^{\bar{\delta}_i} p(\hat{x}_i) d\hat{x}_i \\ &= \int_{\gamma_{ij}}^{\gamma_{ij+1}} \left(\frac{\gamma_{ij+1} - \hat{x}_i}{\gamma_{ij+1} - \gamma_{ij}} \right)^{\bar{\delta}_i} p(\hat{x}_i) d\hat{x}_i \\ &= - \frac{(\gamma_{ij+1} - x_i)^{\bar{\delta}_i} (x_i - \gamma_{ij+1}) (\bar{\delta}_i x_i + x_i + \gamma_{ij+1} - a_i (2 + \bar{\delta}_i))}{(\gamma_{ij+1} - \gamma_{ij})^{\bar{\delta}_i} (a_i - b_i) (\bar{\delta}_i^2 + 3\bar{\delta}_i + 2)} \Bigg|_{x_i=\gamma_{ij}}^{\gamma_{ij+1}} \end{aligned} \tag{A.2}$$

- (2) If $j = J_i$, i.e., $\gamma_i^k = \gamma_{iJ_i}$, by Eq. (7b), there is

$$\begin{aligned} I_i^k(\gamma_i^k) &= \int_{\gamma_{iJ_i-1}}^{\gamma_{iJ_i}} \alpha_{iJ_i}(\hat{x}_i)^{\bar{\delta}_i} p(\hat{x}_i) d\hat{x}_i = \int_{\gamma_{iJ_i-1}}^{\gamma_{iJ_i}} \left(\frac{\hat{x}_i - \gamma_{iJ_i-1}}{\gamma_{iJ_i} - \gamma_{iJ_i-1}} \right)^{\bar{\delta}_i} p(\hat{x}_i) d\hat{x}_i \\ &= - \frac{(\gamma_{iJ_i-1} - x_i)^{\bar{\delta}_i} (x_i - \gamma_{iJ_i-1}) (\bar{\delta}_i x_i + x_i + \gamma_{iJ_i-1} - a_i (2 + \bar{\delta}_i))}{(\gamma_{iJ_i-1} - \gamma_{iJ_i})^{\bar{\delta}_i} (a_i - b_i) (\bar{\delta}_i^2 + 3\bar{\delta}_i + 2)} \Bigg|_{x_i=\gamma_{iJ_i-1}}^{\gamma_{iJ_i}} \end{aligned} \tag{A.3}$$

- (3) If $j = 2, \dots, J_i - 1$, i.e., $\gamma_i^k = \gamma_{ij}$, from Eqs. (7a) and (7b), there is

$$\begin{aligned} I_i^k(\gamma_i^k) &= \int_{\gamma_{ij-1}}^{\gamma_{ij}} \alpha_{ij}(\hat{x}_i)^{\bar{\delta}_i} p(\hat{x}_i) d\hat{x}_i + \int_{\gamma_{ij}}^{\gamma_{ij+1}} \alpha_{ij}(\hat{x}_i)^{\bar{\delta}_i} p(\hat{x}_i) d\hat{x}_i \\ &= \int_{\gamma_{ij-1}}^{\gamma_{ij}} \left(\frac{\hat{x}_i - \gamma_{ij-1}}{\gamma_{ij} - \gamma_{ij-1}} \right)^{\bar{\delta}_i} p(\hat{x}_i) d\hat{x}_i + \int_{\gamma_{ij}}^{\gamma_{ij+1}} \left(\frac{\gamma_{ij+1} - \hat{x}_i}{\gamma_{ij+1} - \gamma_{ij}} \right)^{\bar{\delta}_i} p(\hat{x}_i) d\hat{x}_i, \end{aligned} \tag{A.4}$$

where the analytic result of Eq. (A.4) can be obtained using Eqs. (A.2) and (A.3).

Thus, the statistical utility of a belief rule can be calculated after substituting Eqs. (A.2), (A.3), (A.4) into Eq. (16). Moreover, the statistical utility is put into the proposed sequential adaptive learning algorithm to construct a BRB. Similarly, if the quantitative attribute is a random variable, its statistical utility can also be calculated.

Similarly, we can also suppose that the sampling probability density function $p(\hat{x})$ obeys the other distributions such as normal distribution and exponential distribution.

References

Chen, W. T., & Saif, M. (2005). A novel fuzzy system with dynamic rule base. *IEEE Transactions on Fuzzy Systems*, 13(5), 569–582.

Chung, P. J., & Bohme, J. F. (2005). Recursive EM and SAGE-inspired algorithms with application to DOA estimation. *IEEE Transactions on Signal Processing*, 53(8), 2664–2677.

Dempster, A. P. (1968). A generalization of Bayesian inference. *Journal of the Royal Statistical Society, Series B*, 30(2), 205–247.

Dempster, A. P., Laird, N., & Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society, Series B*, 39(1), 1–38.

Halpern, J. Y., & Fagin, R. (1992). Two views of belief: Belief as generalized probability and belief as evidence. *Artificial Intelligence*, 54(3), 275–317.

Huang, G. B., Sundararajan, N., & Saratchandran, P. (2004). An efficient sequential learning algorithm for growing and pruning RBF (GAP-RBF) networks. *IEEE Transactions on Systems, Man, and Cybernetics-Part B: Cybernetics*, 34(6), 2284–2292.

Huang, G. B., Sundararajan, N., & Saratchandran, P. (2005). A generalized growing and pruning RBF (GGAP-RBF) neural network for function approximation. *IEEE Transactions on Neural Network*, 16(1), 57–67.

Huang, C. L., & Yong, K. (1981). *Multiple attribute decision making methods and applications, a state-of-art survey*. Berlin: Springer-Verlag.

Lewis, C. D. (1982). *Industrial and business forecasting method*. Butterworth Scientific.

- Liu, J., Yang, J. B., & Sii, H. S. (2005). Engineering system safety analysis and synthesis using fuzzy rule based evidential reasoning approach. *Quality and Reliability Engineering International*, 21(4), 387–411.
- Lu, Y. W., Sundararajan, N., & Saratchandran, P. (1997). A sequential learning scheme for function approximation using minimal radial basis function neural networks. *Neural Computation*, 9(2), 461–478.
- Lu, Y. W., Sundararajan, N., & Saratchandran, P. (1998). Performance evaluation of a sequential minimal radial base function (RBF) neural network learning algorithm. *IEEE Transactions on Neural Network*, 9(2), 308–318.
- Rong, H. J., Sundararajan, N., Huang, G. B., & Saratchandran, P. (2006). Sequential adaptive fuzzy inference system (SAFIS) for nonlinear system identification and prediction. *Fuzzy Sets and Systems*, 157(9), 1260–1275.
- Shafer, G. (1976). *A mathematical theory of evidence*. Princeton, NJ: Princeton Univ. Press.
- Sun, R. (1995). Robust reasoning: Integrating rule-based and similarity-based reasoning. *Artificial Intelligence*, 75(2), 241–295.
- Titterton, D. M. (1984). Recursive parameter estimation using incomplete data. *Journal of the Royal Statistical Society, Series B*, 46(2), 257–267.
- Walley, P. (1996). Measures of uncertainty in expert system. *Artificial Intelligence*, 83(1), 1–58.
- 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.
- Xu, Y., Liu, J., Ruan, D., & Li, W. J. (2002). Fuzzy reasoning based on generalized fuzzy if-then rules. *International Journal of Intelligent Systems*, 17(10), 977–1006.
- 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 multi-attribute decision analysis under uncertainties. *European Journal of Operational Research*, 131(1), 31–61.
- Yang, J. B., & Xu, D. L. (1999). *Intelligent decision system via evidential reasoning, version 1.1*. Cheshire, UK: IDSL.
- Yang, J. B., Liu, J., Wang, J., Sii, H. S., & Wang, H. W. (2006). Belief rule-base inference methodology using the evidential reasoning approach – RIMER. *IEEE Transactions on Systems, Man, and Cybernetics – Part A: Systems and Humans*, 36(2), 266–285.
- Yang, J. B., Liu, J., Xu, D. L., Wang, J., & Wang, H. W. (2007). Optimal learning method for training belief rule based systems. *IEEE Transactions on Systems, Man, and Cybernetics – Part A: Systems and Humans*, 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 – Part A: Systems and Humans*, 24(1), 1–18.
- Yang, J. B., Wang, Y. M., Xu, D. L., & Chin, K. S. (2006). The evidential reasoning approach for MADA under both probabilistic and fuzzy uncertainties. *European Journal of Operational Research*, 171(1), 309–343.
- 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(4), 7700–7709.