# WILEY

Check for updates

# A novel nonlinear causal inference approach using vector-based belief rule base

Zhang Zhenjie<sup>1</sup> <sup>©</sup> | Xu Xiaobin<sup>1</sup> <sup>©</sup> | Chen Peng<sup>1</sup> <sup>©</sup> | Wu Xudong<sup>1</sup> <sup>©</sup> | Xu Xiaojian<sup>1</sup> <sup>©</sup> | Wang Guodong<sup>2</sup> <sup>©</sup> | Dustdar Schahram<sup>3</sup> <sup>©</sup>

<sup>1</sup>School of Automation (School of Artificial Intelligence), Hangzhou Dianzi University, Hangzhou, Zhejiang, China <sup>2</sup>Shanghai Institute of Computing Technology, Shanghai, Jingan District, China

<sup>3</sup>Institute of Information Systems Engineering, Vienna University of Technology, Vienna, Austria

#### Correspondence

Xu Xiaobin, School of Automation (School of Artificial Intelligence), Hangzhou Dianzi University, Baiyang Street, Hangzhou, Zhejiang 310018, China. Email: xuxiaobin1980@163.com

#### **Funding information**

NSFC-Zhejiang Joint Fund for the Integration of Industrialization and Informatization, China, Grant/Award Number: U1709215; Zhejiang Province Outstanding Youth Fund, Grant/Award Number: R21F030005; Open Fund of National Engineering Research Centre for Water Transport Safety, China, Grant/Award Number: A2020002; National Youth Science Fund Project of National Natural Science Foundation of China (CN), Grant/Award Number: 61903108; Zhejiang Province Key R&D projects, Grant/Award Number: 2019C03104; 2018C04020; Zhejiang Province Public Welfare Technology Application Research Project, Grant/Award Number: LGF20H270004; LGF19H180018; Key project of Zhejiang Provincial Medical and Health Science and Technology Plan, Grant/Award Number: WKJ-ZJ-2038

#### Abstract

When using the belief rule base (BRB) methodology to deal with the nonlinear causal inference problems, combinatorial explosion often occurs due to overnumbered antecedent attributes, resulting in poor performance. Therefore, this paper proposes a novel nonlinear causal inference approach based on vectorbased BRB. In the modeling process of BRB, the original attributes are ranked by contribution rate and transformed into attribute vectors. Meanwhile, combined with the k-means method, appropriate referential vectors are obtained. Thereby a vectorbased BRB can be established. In the inference process of BRB, the idea of full activation of vector-based rules is presented. By calculating the spatial matching degree of the testing sample and the referential vectors, activation weights of the rules which are used in the evidential reasoning algorithm are acquired. Experimental results of a nonlinear function with four-dimensional input and the pipeline leakage detection data show the effectiveness and superiority of the proposed approach.

5006 WILEY

#### K E Y W O R D S

attribute vector matching, evidential reasoning, full activation, nonlinear causal inference, vector-based belief rule base

### **1** | INTRODUCTION

Belief rule base (BRB) is developed on the traditional If–Then rules. When using the BRB methodology to deal with qualitative and quantitative issues, the uncertainty of information can be maintained because of the introduction of the belief degree.<sup>1,2</sup> Actually, various approaches can be used to handle the uncertainty of information, such as Dempster–Shafer (DS) evidence theory,<sup>3,4</sup> fuzzy sets,<sup>5</sup> rough sets,<sup>6</sup> entropy theory,<sup>7,8</sup> and so on. These approaches have been applied in decision making,<sup>9</sup> reliability analysis,<sup>10</sup> pattern classification,<sup>11</sup> and complex network analysis.<sup>12–14</sup> The BRB inference methodology using the evidential reasoning approach (RIMER) is an important part of the BRB. It effectively combines the advantages of the DS evidence theory,<sup>3,4</sup> fuzzy theory,<sup>5</sup> and traditional If–Then rules.<sup>15</sup> Therefore, it has been widely used to establish nonlinear causal inference models in many engineering fields.<sup>16–24</sup>

The RIMER can be mainly divided into two parts: one is the expression of knowledge and the other is the inference of knowledge.<sup>1,25</sup> The former is to present knowledge through rules in the BRB and the latter is to integrate the activated rules through the evidential reasoning (ER) algorithm.<sup>26</sup> In practical applications, the number of rules in the BRB has become an important factor affecting its performance. Actually, the number of rules is determined by the number of antecedent attributes and the referential values of each attribute. If there are a large number of antecedent attributes or referential values for the attributes, the number of rules will increase exponentially, which is prone to the combinatorial explosion.<sup>27</sup> For example, if the number of antecedent attributes is 5 and each attribute has four referential values, a complicated and large-scale rule base which contains 1024 (i.e., 4<sup>5</sup>) rules will be generated. As a result, it restricts the performance of the BRB to deal with multi-input problems.

Many studies have been undertaken to tackle the aforementioned challenge. These studies can be categorized into two types.

1. Antecedent attributes/referential values reduction (also called structure learning/parameter learning) approaches for an established BRB model. In these studies, the most representative antecedent attributes/referential values were selected and therefore the original BRB model could be downsized. For example, Chang et al. proposed a BRB structure learning method by using the multiple dimensionality reduction techniques, including gray target, multi-dimensional scaling, isomap, and principal component analysis (PCA) to reduce the number of rules.<sup>27</sup> Zhou et al. used the statistical utility to determine which rules were needed to be retained, and thus to construct more compact BRB systems.<sup>28</sup> However, such methods relied on the expert knowledge or extra information heavily. To solve this problem, Wang et al. further explored BRB structure learning using techniques from the rough set theory.<sup>29</sup> To improve the performance of attribute selection, Yang et al. constructed an ensemble BRB system with diverse antecedent attributes selection, in which six different methods (i.e., information gain, gain ratio, chi-squared, ReliefF, OneR classifier, and support vector machine) were applied.<sup>30</sup> Li et al. optimized the referential values by the fuzzy subtractive clustering algorithm.<sup>31</sup> Besides, joint approaches integrating BRB structure learning and

parameter learning were proposed.<sup>32,33</sup> In such approaches, a comprehensive optimization objective called Akaike Information Criterion (AIC) was introduced.<sup>34</sup> However, in these approaches, the redundant rules were reduced after the rule base had been established. When reducing the antecedent attributes or referential values, part of the historical information was also discarded. Consequently, the accuracy of inference results could not be guaranteed.

2. Disjunctive assumption for BRB modeling. The disjunctive BRB modeling approach was different from the traditional BRB which was constructed on the assumption of conjunctive. The disjunctive BRB required far fewer rules in comparison with a conjunctive BRB in the same belief structure (with the same referential values for the same attribute).<sup>35</sup> Chang et al. presented a new rule activation and weight calculation procedures for the disjunctive BRB and applied it in the classification problems.<sup>36</sup> Yang et al. studied the bridge risk assessment problem with an extended disjunctive BRB.<sup>37</sup> Although the disjunctive BRB could help avoid the combinatorial explosion problem, it was difficult to directly obtain a disjunctive BRB from either historical data or experts' knowledge.<sup>35,38</sup>

In short, the above studies have the following limitations. In the first type of approaches, loss of information would occur when reducing the rule number. In the second type of approaches, the interpretability of rules in the disjunctive assumption was needed to be strengthened. Therefore, the problem of "combinatorial explosion" of rules cannot be solved completely. In this paper, a novel vector-based BRB is proposed. It aims to both effectively reduce the scale of the BRB and improve its performance. In the modeling process of BRB, the original attributes are transformed into attribute vectors to reduce the rule number. In the inference process of BRB, the original BRB activation method is extended from one-dimensional to multidimensional vectors to improve the accuracy.

The remainder of this paper is structured as follows: Section 2 introduces the basic knowledge of the BRB inference; Section 3 presents the BRB inference model using attribute vector activation; Section 4 discusses the experimental analysis and performance comparison and Section 5 is the conclusion.

# **2** | THE BRB FOR INFERENCE

The BRB model is designed on the basis of the belief structure. A rule is used to describe causal relationships as well as uncertainty between antecedent attributes and their associated consequent attribute.<sup>1</sup>

#### 2.1 | Representation of the BRB

In the RIMER, the *k*th rule  $Rule_k$  is represented as

$$Rule_{k}: \text{ If } x_{1} \text{ is } A_{1}^{k} \wedge x_{2} \text{ is } A_{2}^{k} \wedge \dots \wedge x_{T_{k}} \text{ is } A_{T_{k}}^{k}$$

$$Then \ \{(D_{1}, \beta_{1,k}), (D_{2}, \beta_{2,k}), ..., (D_{N}, \beta_{N,k})\},$$
(1)

where  $x_i$   $(i = 1, 2, ..., T_k)$  is the antecedent attribute,  $T_k$  is the number of the antecedent attributes;  $A_i^k$  is the referential value of the *i*th antecedent attribute in the *k*th rule;  $A_i = \{A_{i,j}\}$ 

 $j = 1, 2, ..., v_i$ ,  $A_{i,j}$  is the *j*th referential value of the *i*th antecedent attribute  $x_i, v_i$  is the number of referential values of the *i*th antecedent attribute;  $D_n$  (n = 1, 2, ..., N) is the referential value of the consequent attribute, N is the number of referential values of the consequent attribute;  $\beta_{n,k}$  is the belief degree of the referential value of consequent attribute in the *k*th rule.

#### 2.2 | Inference process of the BRB

BRB inference is mainly divided into two parts: calculation of activation weight and outputs fusion based on ER algorithm.

# 2.2.1 | Calculation of activation weight

The activation weight is mainly affected by the antecedent attribute, rule weight, and input data. Therefore, the activation weight of the *k*th rule  $\omega_k$  can be calculated as follows:

$$\omega_{k} = \frac{\theta_{k} \prod_{i=1}^{T_{k}} (\alpha_{i}^{k})^{\delta_{i}}}{\sum_{k=1}^{RN} (\theta_{k} \prod_{i=1}^{T_{k}} (\alpha_{i}^{k})^{\overline{\delta}_{i}})},$$
(2)

$$\bar{\delta}_i = \frac{\delta_i}{\max_{i=1,2,\dots,T_k} \{\delta_i\}},\tag{3}$$

where  $\delta_i$  is the weight of the *i*th antecedent attribute,  $\theta_k$  (k = 1, 2, ..., RN) is the rule weight, *RN* is the number of rules, and  $\alpha_i^k$  is the matching degree of the referential value of the *i*th antecedent attribute in the *k*th rule which can be calculated as follows:

$$\alpha_{i,j}^{k} = \frac{A_{i,j+1}^{k} - x_{i}}{A_{i,j+1}^{k} - A_{i,j}^{k}}, \quad A_{i,j}^{k} \le x_{i} \le A_{i,j+1}^{k}, \tag{4}$$

$$\alpha_{i,j+1}^{k} = \frac{x_i - A_{i,j}^{k}}{A_{i,j+1}^{k} - A_{i,j}^{k}}, \quad A_{i,j}^{k} \le x_i \le A_{i,j+1}^{k}, \tag{5}$$

$$\alpha_{i,t}^k = 0, \quad t = 1, ..., v_i, \quad t \neq j, j + 1,$$
(6)

where  $v_i$  and  $\alpha_{i,j}^k$  are the number of referential values and the matching degree of the *j*th referential value of the *i*th antecedent attribute in the *k*th rule, respectively.

#### 2.2.2 | Outputs fusion based on ER algorithm

Suppose the weight of each rule is  $\theta_k$ , the initial belief degree of corresponding referential level determined by expert knowledge is  $\beta_n^k$ , then the output of the BRB model  $\hat{O}$  can be obtained by the *RN* rules combined according to the antecedent attributes. The analytical format of the ER algorithm is as follows:

$$\hat{O} = \{ (D_n, \hat{\beta}_n), n = 1, 2, ..., N \},$$
(7)

/II FY

$$\hat{\beta}_n = \frac{\mu \times \left[\prod_k^{RN} \left(\omega_k \beta_n^k + 1 - \omega_k \sum_{n=1}^N \beta_n^k\right) - \prod_k^{RN} \left(1 - \omega_k \sum_{n=1}^N \beta_n^k\right)\right]}{1 - \mu \times \left[\prod_k^{RN} (1 - \omega_k)\right]},\tag{8}$$

$$\mu = \left[\sum_{n=1}^{N} \prod_{k=1}^{RN} \left(\omega_k \beta_n^k + 1 - \omega_k \sum_{n=1}^{N} \beta_n^k\right) - (N-1) \prod_{k=1}^{RN} \left(1 - \omega_k \sum_{n=1}^{N} \beta_n^k\right)\right]^{-1}, \quad (9)$$

where  $\hat{\beta}_n$  (n = 1, 2, ..., N) is the belief degree of the output referential value of  $D_n$  after fusion. Then, the final outcomes can be generated as follows:

$$y = \sum_{n=1}^{N} (D_n \times \hat{\beta}_n).$$
(10)

5009

WILEY

# 3 | ATTRIBUTE VECTOR-BASED BRB INFERENCE MODEL

For a set of inputs  $X = (x_1, ..., x_r, ..., x_{T_k})$ , where  $T_k$  is the number of attributes of each input, the purpose of BRB modeling is to obtain the antecedent part of the BRB by crossmatching the referential values of each antecedent attribute. Therefore, the dimension of the antecedent attributes and the number of referential values of each attribute will directly affect the scale of the BRB. If the number of attributes  $T_k$  is too large, the scale of the established BRB will increase. Meanwhile, the number of parameters that need to be determined and optimized will also increase. The nonlinear causal inference approach based on vector-based BRB (as shown in Figure 1) proposed in this paper can effectively solve the problem.

The proposed approach mainly includes the following parts: (1) *Generation of attribute vectors*. Obtaining the importance of the  $T_k$  original attributes by calculating the contribution rate of the principal component, and ranking the attributes according to the results. Then, dividing the ranked  $T_k$  attributes into J attribute vectors. (2) *Calculation of attribute vector referential matrix*. Obtaining the referential vectors based on the attribute vectors sample statistics. (3) *Full activation of attribute vector-based rules*. Calculating the spatial matching degree of the training sample to the referential vectors to get the activation weights of the rules and using the ER algorithm to obtain the final outcomes. (4) *Model optimization*. Obtaining the optimal parameters by training sample data.

### 3.1 | Generation of attribute vectors

As a multivariate statistical technology, PCA methodology can realize multidimensional orthogonal linear transformation based on statistical features.<sup>39</sup> It is often used for feature extraction of data and thus has a wide range of applications in the fields of pattern recognition and image processing. The new variables (principal component variables) obtained by the PCA methodology are a linear combination of the original variables. Since the principal components are arranged in the order of contribution rate from largest to smallest, it can retain important features in high-dimensional data, remove noise and unimportant features, and thus achieve the purpose of improving data processing speed. Therefore, in this section, we use the principle of "ranking by contribution rate" of the principal components in PCA to realize the importance



**FIGURE 1** Nonlinear causal inference approach based on vector-based BRB. BRB, belief rule base; ER, evidential reasoning; PCA, principal component analysis; RN, rule number; SLP, sequential linear programming [Color figure can be viewed at wileyonlinelibrary.com]

ranking of the original attributes, and obtain *J* attribute subvectors with different importance according to the ranking results. The specific steps are as follows:

Step 1. For an input data sample set  $X = (x_1, ..., x_r, ..., x_{T_k})$ ,  $x_r = [x_r(1), ..., x_r(t), ..., x_r(T)]^T$ , where  $r = 1, 2, ..., T_k$ ,  $x_r$  is a vector composed of *T* sampling values of the *r*th attribute variable, while the superscript "T" represents the transpose of the matrix. Using the following formula to map the  $T_k$  original attributes to the principal component space.

$$\begin{cases}
F_1 = w_{11}x_1 + w_{21}x_2 + \dots + w_{T_k1}x_{T_k}, \\
F_2 = w_{12}x_1 + w_{22}x_2 + \dots + w_{T_k2}x_{T_k}, \\
\vdots \\
F_i = w_{1i}x_1 + w_{2i}x_2 + \dots + w_{T_ki}x_{T_k}, \\
\vdots \\
F_{T_k} = w_{1T_k}x_1 + w_{2T_k}x_2 + \dots + w_{T_kT_k}x_{T_k}.
\end{cases}$$
(11)

For the *i*th principal component  $F_i = w_{1i}x_1 + w_{2i}x_2 + \cdots + w_{ri}x_r + \cdots + w_{T_ki}x_{T_k}$ , it is a *T*-dimensional vector since  $x_r$  is *T*-dimensional;  $w_{ri}$  is the weight coefficient of the *r*th attribute for the *i*th principal component. The principal component satisfies the following conditions: first,  $F_i$  and  $F_r$  are irrelevant ( $i \neq r$ ;  $i, r = 1, 2, ..., T_k$ ); second, the variance of  $F_i$  is greater than that of  $F_r$  (i < r), that is, the variance is ranked in descending order. The transformation matrix *W* formed by weight coefficients is as follows:

$$W = \begin{bmatrix} w_{11} & w_{12} & \cdots & w_{1i} & \cdots & w_{1T_k} \\ w_{21} & w_{22} & \cdots & w_{2i} & \cdots & w_{2T_k} \\ \vdots & \vdots & \vdots & \vdots \\ w_{T_k1} & w_{T_k2} & \cdots & w_{T_ki} & \cdots & w_{T_kT_k} \end{bmatrix}.$$
 (12)

The principal component can be generated by  $F = [F_1, F_2, ..., F_{T_k}] = XW$ .

Step 2. For the input data set X, it needs to be centralized:

$$\tilde{x}_r(t) = x_r(t) - \tau(r), \tag{13}$$

$$\tau(r) = \frac{1}{T} \sum_{t=1}^{T} x_r(t),$$
(14)

where  $\tau(r)$  is the mean value of attribute  $x_r$ .

After the data centralization, a new matrix can be obtained as

$$\tilde{X} = \begin{bmatrix} \tilde{x}_{1}(1) & \tilde{x}_{2}(1) & \cdots & \tilde{x}_{r}(1) & \cdots & \tilde{x}_{T_{k}}(1) \\ \tilde{x}_{1}(2) & \tilde{x}_{2}(2) & \cdots & \tilde{x}_{r}(2) & \cdots & \tilde{x}_{T_{k}}(2) \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \tilde{x}_{1}(t) & \tilde{x}_{2}(t) & \cdots & \tilde{x}_{r}(t) & \cdots & \tilde{x}_{T_{k}}(t) \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \tilde{x}_{1}(T) & \tilde{x}_{2}(T) & \cdots & \tilde{x}_{r}(T) & \cdots & \tilde{x}_{T_{k}}(T) \end{bmatrix}.$$
(15)

Then the correlation coefficient matrix V can be generated by

$$V = \frac{1}{T-1} \tilde{X}^{\mathrm{T}} \tilde{X}.$$
 (16)

Carrying out eigenvalue decomposition on the matrix *V*, the eigenvector corresponding to each eigenvalue  $\lambda_r$  ( $r = 1, 2, ..., T_k$ ) can be obtained, and then the contribution rate of the *r*th principal component  $F_r$  can be calculated by

$$\varphi_r = \frac{\lambda_r}{\sum_{r=1}^{T_k} \lambda_r}.$$
(17)

Step 3. Ranking the  $T_k$  original attributes according to contribution rate  $\varphi_r$ , and resetting the number as  $x'_1, x'_2, ..., x'_{T_k}$ , then dividing the new ranked  $T_k$  attributes into J vectors.

$$\begin{cases} R_{1} = [x'_{1}, x'_{2}, ..., x'_{h}], \\ R_{2} = [x'_{h+1}, x'_{h+2}, ..., x'_{2h}], \\ \vdots \\ R_{r} = [x'_{(r-1)h+1}, x'_{(r-1)h+2}, ..., x'_{rh}], \\ \vdots \\ R_{J} = [x'_{(J-1)h+1}, x'_{(J-1)h+2}, ..., x'_{T_{k}}], \end{cases}$$

$$(18)$$

where J represents the number of obtained attribute vectors, and h represents the number of original attributes contained in each vector.

### 3.2 | Calculation of attribute vector referential matrix

The referential values are the most representative points which are usually determined based on experts' experience. Since the inputs of the BRB become multidimensional vectors, the original referential values which are numerical type will no longer be applicable. Therefore, the

WILFY

referential vector is introduced. However, determining suitable referential vectors manually is a challenging task. To improve the objectivity and effectiveness of the selected referential vectors, a new solution based on data analysis method is provided. As a mature data analysis method, *k*-means clustering method is used. For the combined *J* attribute vectors in the formula (18), dividing each dimension of sample data in each attribute vector  $x_i = [x_i(1), ..., x_r(j), ..., x_i(T)]$  into  $K_i$  clusters  $\{C_{1,i}, ..., C_{k,i}, ..., C_{K_i,i}\}$  in ascending order, where  $K_i$  is the number of cluster centers corresponding to the *i*th group of data. The steps are as follows:

Step 1. Randomly selecting  $K_i$  data  $\{x_i^1, x_i^2, ..., x_i^{K_i}\}$  as the centers of clusters  $\{C_{1,i}, ..., C_{k,i}, ..., C_{K_i,i}\}$ .

Step 2. Calculating the distance  $\{d_1(i), d_2(i), ..., d_{K_i}(i)\}$  of each data sample (except the cluster center) and each cluster center, and assigning it to the cluster in which the distance is smallest.

*Step* 3. Recalculating the center of each cluster according to formula (19), where  $n(C_{k,i})$  represents the number of elements in the cluster  $C_{k,i}$ .

$$x_{i}^{r} = \frac{\sum x_{i}(r)}{n(C_{k,i})}.$$
(19)

Step 4. Repeating the above steps until the cluster centers no longer change.

It should be noted that the number of cluster centers  $K_i$  needs to be determined according to the elements in the vector  $R_r$  (r = 1, 2, ..., J) combined in Section 3.1. For example, if  $R_1 = (x_1, x_2)$ , then the number of cluster centers  $K_1$  and  $K_2$  of  $x_1$  and  $x_2$  should be equal. Then finding the cluster centers of  $R_r$  to determine the corresponding referential vector. To ensure the diversity of samples, the referential vector must cover the range of all sample data, that is, the interval of the referential vector should be greater than or equal to the interval of its corresponding sample. For this reason, the upper and lower bounds of the sample space are used as the upper and lower bounds of its referential vector. For example, the upper and lower bounds of  $x_1$  and  $x_2$  are, respectively,  $x_1^{\min}, x_1^{\max}, x_2^{\min}, x_2^{\max}$ , and the cluster centers are  $x_1^1, x_1^2, ..., x_1^{K_1}$  and  $x_2^1, x_2^2, ..., x_2^{K_2}$ , respectively, where  $K_1 = K_2$ . Then the referential vector of  $R_1$  is expressed as  $A_1 = (x_1^{\min} x_2^{\min}, x_1^{1} x_2^1, ..., x_1^{K_1} x_2^{K_2}, x_1^{\max} x_2^{\max})$ .

Similarly, the referential values of output  $D = (D_1, D_2, ..., D_N)$  can be obtained, where N is the number of output data referential values,  $D_1$  and  $D_N$  are the lower and upper bounds, respectively,  $\{D_2, ..., D_{N-1}\}$  are the cluster centers obtained by the *k*-means clustering method.

### 3.3 | Full activation of attribute vector-based rules

According to the formulas (4)–(6), only the matching degree between the input data and its two closest referential values is calculated. However, the attribute vector cannot be cast into a specific interval, like, the numerical input. In this case, if only the reference vectors closest to the attribute vector are considered, it is difficult to ensure the accuracy of the matching degree. Therefore, the full activation method is introduced to calculate the similarity with all reference vectors.

Since the input vector  $R_r$  and its referential vector  $A_r$  are obtained in Section 3.2, the distance of  $R_r$  and  $A_r$  can be calculated by

$$d_r^{K_r} = |R_r^{K_r} - A_r^{K_r}|, (20)$$

 $\mathcal{N}$ II FY

$$d_r^{K_r} = \begin{cases} \left[ \left( R_r^1 - A_r^1 \right)^2 + \left( R_r^2 - A_r^2 \right)^2 + \dots + \left( R_r^h - A_r^h \right)^2 \right]^{1/h} & \text{if } c = h, \\ \left[ \left( R_r^1 - A_r^1 \right)^2 + \left( R_r^2 - A_r^2 \right)^2 + \dots + \left( R_r^c - A_r^h \right)^2 \right]^{1/c} & \text{if } c < h, \end{cases}$$
(21)

where c is the number of original attributes in the last attribute vector.

Since the distance and belief degree are inversely related, the distance needs to be reciprocated and normalized. After that, the matching degree  $\alpha_r^{K_r}$  can be obtained as follows:

$$\begin{cases} g_r^{K_r} = \frac{1}{d_r^{K_r}}, \quad d_r^{K_r} \neq 0, \\ \alpha_r^{K_r} = \frac{g_r^{K_r}}{\sum_{r=1}^{K_f} g_r^{K_r}}, \end{cases}$$
(22)

5013

WILE

$$\begin{cases} d_r^{K_r} = 0, \\ \alpha_r^{K_r} = 1, \quad \text{else} = 0. \end{cases}$$
(23)

#### 3.4 | Optimal training

Then the outputs of the BRB can be obtained according to the formulas (2), (3), and (7)–(10) in Section 2. Since some parameters of the initial BRB are determined by experts based on their own experience and professional knowledge, deviations and misjudgments will inevitably occur if the problem is too complex. It will affect the accuracy of the BRB model. To improve the performance of the BRB model, historical sample data can be used to train the parameters, that is, to optimize the antecedent attribute weights, rule weights, and the belief degree of output reference values.

Figure 2 shows the BRB optimization model, where *X* is the input sample data set, *O* is the observation value of the actual sample set,  $\hat{O}$  is the actual output value of the BRB model,  $P = \{\beta_n^k, \theta_k, \delta_i \mid n = 1, 2, ..., N; i = 1, 2, ..., M; k = 1, 2, ..., RN\}$  the parameter set to be optimized,  $\xi(P)$  is the error between *O* and  $\hat{O}_i$ .

**Objective function:** 

$$\begin{cases} \min\{\xi(P)\}, \\ \xi(P) = \frac{1}{n} \sum_{i=1}^{n} (O - \hat{O})^2. \end{cases}$$
(24)

Restrictions:

$$\begin{cases} 0 \le \beta_n^k \le 1, & n = 1, 2, ..., N, & k = 1, 2, ..., RN, \\ 0 \le \theta_k \le 1, & k = 1, 2, ..., RN, \\ 0 \le \delta_i \le 1, & i = 1, 2, ..., T_k. \end{cases}$$
(25)



FIGURE 2 BRB optimization model. BRB, belief rule base

It is a nonlinear multi-objective optimization problem. The sequential linear programming (SLP) optimization algorithm is used to solve it. The basic idea of SLP is to transform the nonlinear programming problem into a series of linear programming problems, then use the simplex method to obtain the solution set of these linear programming problems, and finally obtain the solution of the original nonlinear programming step by step.

The antecedent attribute weights, rule weights, and belief degree of the output referential value obtained after optimization are the parameters of the final BRB model.

On the basis of the above steps, the rule number as well as the optimized parameter number can be reduced.

**Theorem 1.** Let  $M_p$  (p = 1, 2, ..., J) be the number of the referential vectors of attribute vector  $R_p$ , J be the number of attribute vectors, h be the number of original attributes contained in each vector,  $T_k$  be the number of attributes of each input,  $RN_1$  and  $RN_2$  be the rule number in original BRB and vector-based BRB, respectively,  $OPN_1$  and  $OPN_2$  be the optimized parameter number in original BRB and vector-based BRB and vector-based BRB, respectively, N be the number of referential values of the consequent attribute, then the rule number can be reduced at least  $\prod_{p=1}^{J-1} (M_p)^{h-1}$  times, the optimized parameter number can also be reduced at least  $\prod_{p=1}^{J-1} (M_p)^{h-1}$  times.

*Proof.* According to Sections 3.1 and 3.2, the number of the referential values of each attribute in  $R_p$  is equal to  $M_p$ . Then, the  $RN_1$  and  $RN_2$  can be expressed as  $RN_1 = (M_f)^{T_k - (J-1)h} \cdot \prod_{p=1}^{J-1} (M_p)^h$  and  $RN_2 = \prod_{p=1}^J M_p$ , respectively. Defined  $C_{RN} = (M_f)^{T_k - (J-1)h-1} \cdot \prod_{p=1}^{J-1} (M_p)^{h-1}$  as the rule reduction coefficient, then  $RN_2 = \frac{1}{C_{RN}} \cdot RN_1$ . According to formula (18), there exists  $(J-1)h + 1 \leq T_k \leq Jh$ , then  $\prod_{p=1}^{J-1} (M_p)^{h-1} \leq C_{RN} \leq \prod_{p=1}^J (M_p)^{h-1}$ . Therefore, the rule number can be reduced at least  $\prod_{p=1}^{J-1} (M_p)^{h-1}$  times.

Since the  $OPN_1$  and  $OPN_2$  can be expressed as  $OPN_1 = RN_1 \cdot N + RN_1 + T_k$  and  $OPN_2 = RN_2 \cdot N + RN_2 + J$ , respectively. Then  $OPN_2 \approx \frac{1}{C_{RN}} \cdot OPN_1$ . Thus, the optimized parameter number can also be reduced at least  $\prod_{p=1}^{J-1} (M_p)^{h-1}$  times (approximate value).

# 4 | EXPERIMENTAL ANALYSIS AND PERFORMANCE COMPARISON

In this section, the effectiveness and superiority of the proposed approach are verified through two cases. One is a simulation data case study, and the other is a practical data case study about pipeline leakage detection. For ease of explanation, the vector-based BRB approach proposed in this paper is referred to as V-BRB; the original BRB, that is, without modification to the inputs and outputs, is referred to as O-BRB.

The experimental environment configuration is Dell/windows10/4G memory/Intel i5/ matlab2016a.

To evaluate the performance of the algorithm, four indicators, namely, the rule number (*RN*), optimized parameter number (*OPN*), mean square error (*MSE*) between estimated values and actual values, and *AIC* are introduced.

- 1. The *RN* reflects the scale of the system. The smaller the value, the more concise the model.
- 2. The smaller the OPN, the better the model performance.
- 3. The *MSE* between the estimated values and the actual values reflects the accuracy of the inference result. The smaller the value, the better the inference result.
- 4. *AIC* is an index that comprehensively considers the accuracy and complexity of the model. It can reflect the degree of similarity between the established model and the real system. The smaller the *AIC*, the higher the degree of similarity between the established model and the real system. Its calculation formula is as follows:

$$AIC_{BRB} = Pln(P \cdot MSE) + 2OPN, \qquad (26)$$

where P represents the number of data in the data set.

### 4.1 | Case study 1: simulation data

The simulation data are generated by the following nonlinear function:

$$y = 8\sin x_1 + 0.7x_2x_3 - 0.1x_4^4.$$
<sup>(27)</sup>

According to the formula (27), the performance of the proposed approach on the mapping relationship between the input data  $x_i$  and the output data y needs to be verified where the value range of the independent variable  $x_i$  (i = 1, 2, 3, 4) is [0.17, 0.88], [2, 4], [1, 7], and [0, 2]. According to the range of  $x_i$ , the range of y can be calculated as in References [3,35]. Randomly selecting 100 groups of data for the  $x_i$  and the corresponding output y as the sample data set  $U = \{x_1, x_2, x_3, x_4, y\}$ . In the BRB model, the variables  $x_1, x_2, x_3$ , and  $x_4$  are regarded as the antecedent attributes and the variable y is regarded as the consequent attribute.

#### 4.1.1 | Numerical study of V-BRB

Sixty groups in the sample data are selected randomly and set as the training samples, the remaining 40 groups are set as the testing samples.

First, the PCA methodology is used to obtain the weights of the input  $x_i$ . The corresponding values are  $\varphi(x_1) = 0.13$ ,  $\varphi(x_2) = 0.32$ ,  $\varphi(x_3) = 0.43$ , and  $\varphi(x_4) = 0.12$ . Obviously,  $\varphi(x_3) > \varphi(x_2) > \varphi(x_1) > \varphi(x_4)$ . Therefore, the inputs  $x_3$  and  $x_2$  can be combined into a vector  $R_1 = [x_3x_2]$ ,  $x_1$  and  $x_4$  can be combined into a vector  $R_2 = [x_1x_4]$ . Then the four-dimensional input can be converted to two-dimensional. In this case, eight-dimensional referential values are used in the inputs  $R_1$  and  $R_2$ . The referential values are described by fuzzy semantic format. They are limit small (*LS*), very small (*VS*), positive small (*PS*), medium (*M*), positive medium (*PM*), large (*L*), medium large (*VL*). Also, the output is described by six fuzzy semantic values: very small (*VS*), positive small (*PS*), medium (*M*), positive medium (*PM*), large (*L*), and very large (*VL*).

Second, the *k*-means clustering method is used to calculate the input referential values which are (0.17, 0.37, 0.44, 0.53, 0.60, 0.61, 0.63, 0.88), (2, 2.6, 2.8, 3.0, 3.1, 3.3, 3.6, 4), (1, 1.7, 3.8, 3.6, 4), (1, 1.7, 3.8, 3.6, 4), (1, 1.7, 3.8, 3.6, 4), (1, 1.7, 3.8, 3.6, 4), (1, 1.7, 3.8, 3.6, 4), (1, 1.7, 3.8, 3.6, 4), (1, 1.7, 3.8, 3.6, 4), (1, 1.7, 3.8, 3.6, 4), (1, 1.7, 3.8, 3.6, 4), (1, 1.7, 3.8, 3.6, 4), (1, 1.7, 3.8, 3.6, 4), (1, 1.7, 3.8, 3.6, 4), (1, 1.7, 3.8, 3.6, 4), (1, 1.7, 3.8, 3.6, 4), (1, 1.7, 3.8, 3.6, 4), (1, 1.7, 3.8, 3.6, 4), (1, 1.7, 3.8, 3.6, 4), (1, 1.7, 3.8, 3.6, 4), (1, 1.7, 3.8, 3.6, 4), (1, 1.7, 3.8, 4



#### TABLE 1 Referential vectors of inputs

	Semantic value							
Referential vector	LS	VS	PS	M	PM	L	ML	VL
$A_1$	[1 2]	[1.7 2.6]	[3.3 2.8]	[3.5 3.0]	[5.3 3.1]	[5.5 3.3]	[6.4 3.6]	[7 4]
$A_2$	[0.17 0]	[0.4 0.64]	[0.44 1.0]	[0.53 1.1]	[0.60 1.2]	[0.61 1.3]	[0.63 1.5]	[0.88 2]

Abbreviations: *L*, large; *LS*, limit small; *M*, medium; *ML*, medium large; PM, positive medium; *PS*, positive small; *VL*, very large; *VS*, very small.

TABLE 2 Referential values of outputs

Semantic value						
Referential value	VS	PS	M	PM	L	VL
D	3	6.6	10.4	14.0	18.8	22

Abbreviations: L, large; M, medium; PM, positive medium; PS, positive small; VL, very large; VS, very small.

3.5, 5.3, 5.5, 6.4, 7), and (0, 0.64, 1.0, 1.1, 1.2, 1.3, 1.5, 2). Then the referential vectors can be obtained as  $A_1 = ([1 \ 2]; [1.7 \ 2.6]; [3.3 \ 2.8]; [3.5 \ 3.0]; [5.3 \ 3.1]; [5.5 \ 3.3]; [6.4 \ 3.6]; [7 \ 4])$  and  $A_2 = ([0.17 \ 0]; [0.4 \ 0.64]; [0.44 \ 1.0]; [0.53 \ 1.1]; [0.60 \ 1.2]; [0.61 \ 1.3]; [0.63 \ 1.5]; [0.88 \ 2])$ , as shown in Table 1. Similarly, using the *k*-means clustering method to calculate the output referential value D = (3, 6.8, 10.6, 14.4, 19, 22), as shown in Table 2.

It can be seen from Table 1, there are totally 64 (i.e.,  $8 \times 8 = 64$ ) rules that need to be established. The initial BRB is established based on expert knowledge and historical data, as shown in Table 3. The initial antecedent attribute weight  $\delta_i$  and rule weight  $\theta_k$  are both set to 1.

Then, 60 groups of data are input to the initial V-BRB model. The inference effect is shown in Figure 3. Obviously, the performance is poor.

Therefore, optimization training is used to obtain optimal rule weights, antecedent attribute weights, and belief degree. The optimized antecedent attribute weights are 0.8128 and 0.9445, respectively. The optimized V-BRB is shown in Table 4.

The optimized V-BRB model is used to test the 60 groups of training data and 40 groups of test data, respectively, and the inference effect (i.e., the comparison between the estimated values and the true values) is shown in Figures 4 and 5. It can be seen that the inference effect is relatively ideal, indicating that the vector-based BRB reasoning approach proposed in this paper is effective.

#### 4.1.2 | Further analysis of effectiveness

To further illustrate the effectiveness of the V-BRB model in processing the casual inference problem of multidimensional inputs function, the O-BRB model is used to make comparison. It should be noted that the O-BRB model is built on the same 60 groups of data. The experimental results are shown in Table 5.

As shown in Table 5, the indicator RN of the V-BRB model is much better than that of the O-BRB model. In this case, since the four antecedent attributes are transformed into two vectors, there are only 64 rules generating in the V-BRB model. Therefore, 75% of rules can be reduced. It greatly reduces the size of the rule base. When comparing the indicator MSE, the result of the O-BRB model is slightly better than that of the V-BRB model. This is because

TABLE 3 Initial V-BRB

No.	$\theta_k$	$X_1$ and $X_2$	Belief structure
1	1	LS and LS	$\{(D_1, 0), (D_2, 0.9926), (D_3, 0.0074), (D_4, 0), (D_5, 0), (D_6, 0)\}$
2	1	LS and VS	$\{(D_1,0),(D_2,0.9926),(D_3,0.0074),(D_4,0),(D_5,0),(D_6,0)\}$
3	1	LS and PS	$\{(D_1,0),(D_2,0.9926),(D_3,0.0074),(D_4,0),(D_5,0),(D_6,0)\}$
4	1	LS and M	$\{(D_1,0),(D_2,0.9926),(D_3,0.0074),(D_4,0),(D_5,0),(D_6,0)\}$
5	1	LS and PM	$\{(D_1, 0), (D_2, 0.9926), (D_3, 0.0074), (D_4, 0), (D_5, 0), (D_6, 0)\}$
6	1	LS and L	$\{(D_1, 0), (D_2, 0.9926), (D_3, 0.0074), (D_4, 0), (D_5, 0), (D_6, 0)\}$
7	1	LS and ML	$\{(D_1, 0.0846), (D_2, 0.9154), (D_3, 0), (D_4, 0), (D_5, 0), (D_6, 0)\}$
8	1	LS and VL	$\{(D_1, 0.0846), (D_2, 0.9154), (D_3, 0), (D_4, 0), (D_5, 0), (D_6, 0)\}$
36	1	<i>PM</i> and <i>M</i>	$\{(D_1, 0), (D_2, 0), (D_3, 0.5136), (D_4, 0.4864), (D_5, 4), (D_6, 4)\}$
37	1	PM and PM	$\{(D_1,0),(D_2,0),(D_3,0.3871),(D_4,0.6129),(D_5,0),(D_6,0)\}$
38	1	<i>PM</i> and <i>L</i>	$\{(D_1,0),(D_2,0),(D_3,0.7314),(D_4,0.6129),(D_5,0),(D_6,0)\}$
39	1	PM and ML	$\{(D_1,0),(D_2,0),(D_3,0.7314),(D_4,0.6129),(D_5,0),(D_6,0)\}$
40	1	PM and VL	$\{(D_1, 0), (D_2, 0), (D_3, 0.8589), (D_4, 0.1411), (D_5, 0), (D_6, 0)\}$
41	1	L and LS	$\{(D_1,0),(D_2,0),(D_3,0.2880),(D_4,0.7120),(D_5,0),(D_6,0)\}$
42	1	L and VS	$\{(D_1,0),(D_2,0),(D_3,0.2880),(D_4,0.7120),(D_5,0),(D_6,0)\}$
43	1	L and PS	$\{(D_1,0),(D_2,0),(D_3,0.2880),(D_4,0.7120),(D_5,0),(D_6,0)\}$
58	1	VL and VS	$\{(D_1, 0), (D_2, 0), (D_3, 0), (D_4, 0), (D_5, 0.2733), (D_6, 0.7267)\}$
59	1	VL and PS	$\{(D_1,0),(D_2,0),(D_3,0),(D_4,0),(D_5,0.2733),(D_6,0.7267)\}$
60	1	VL and $M$	$\{(D_1, 0), (D_2, 0), (D_3, 0), (D_4, 0), (D_5, 0.6290), (D_6, 0.3710)\}$
61	1	VL and PM	$\{(D_1, 0), (D_2, 0), (D_3, 0), (D_4, 0), (D_5, 0.6290), (D_6, 0.3710)\}$
62	1	VL and L	$\{(D_1, 0), (D_2, 0), (D_3, 0), (D_4, 0), (D_5, 0.5516), (D_6, 0.4484)\}$
63	1	VL and ML	$\{(D_1, 0), (D_2, 0), (D_3, 0), (D_4, 0), (D_5, 0.5516), (D_6, 0.4484)\}$
64	1	VL and VL	$\{(D_1,0),(D_2,0),(D_3,0),(D_4,0),(D_5,0.5516),(D_6,0.4484)\}$

Abbreviations: L, large; LS, limit small; M, medium; ML, medium large; PM, positive medium; PS, positive small; V-BRB, vector-based belief rule base; VL, very large; VS, very small.

the V-BRB model reduces the rules, which will affect the accuracy of the model to a certain extent. Nevertheless, it is acceptable. For the indicator AIC, the value of V-BRB is much smaller than the value of O-BRB, which shows that V-BRB is much closer to the real system. Actually, in the O-BRB model, there are 1796 optimization parameters, while in the V-BRB model only contains 450 optimization parameters. It significantly reduces the complexity of the optimization process.

In conclusion, the V-BRB model proposed in this paper is effective in dealing with the casual inference problem of multidimensional input nonlinear function. Moreover, compared with the traditional model, it has obvious advantages.

WILFY



FIGURE 3 Inference effect of training data of initial vector-based belief rule base [Color figure can be viewed at wileyonlinelibrary.com]

#### 4.2 Case study 2: pipeline leakage detection data

In this section, a more complex practical application about pipeline leakage detection is studied to make further analysis of the proposed approach.

#### Introduction of the data set 4.2.1

The data set comes from a pipeline more than 100 km in length installed in the UK. The pipeline is equipped with flow sensors at the inlet and outlet to detect the oil flow value, and 10 pressure sensors are distributed at the inlet and outlet of the pipeline and in the middle of the pipeline to detect pressure changes. Changes in data reveal the statuses of the pipeline (i.e., leak condition or normal condition). Under normal operations, when inlet flow is larger (or less) than outlet flow, the pressure in the pipeline will build up (or decrease) because the total content in the pipeline is increasing (or decreasing, respectively). However, if the pattern is violated, for example, if the inlet flow is larger than the outlet flow, yet the pressure in the line still decreases, then it is highly likely that there is a leak in the pipeline.

The pipeline is mostly operated in normal condition. To achieve the experimental effect, a valve is used in the middle of the pipeline to control the oil flowing out to achieve the effect of simulating leakage. Each leak lasts for several hours, and the size of the leak is also controlled by the value. The inlet and outlet flow values and pressure values are collected  $(f_0, f_1; p_0, p_1, ..., p_9, p_1, ..., p_9, p_1, ..., p_9)$ respectively). During the leak trial (from #07:00:08# to #12:34:27#), 2008 samples which leak size is 25% were collected at the rate of 10 s per sample, as shown in Figure 6. The 25% mean there were 10,000 L of oil in the pipeline at a certain moment, and 250 L of oil had leaked.

#### 4.2.2 Analysis of effectiveness and superiority

To realize pipeline leakage detection and leak size estimation, the difference between the inlet flow and outlet flow (i.e.,  $f_1 - f_0$ ) and the value of the nonadjacent pressure sensors (i.e., p0, p2,

TABLE 4 Optimized V-BRB

No.	$\theta_k$	$X_1$ and $X_2$	Belief structure
1	0.8364	LS and LS	$\{(D_1, 0.3278), (D_2, 0.6070), (D_3, 0), (D_4, 0), (D_5, 0.0305), (D_6, 0.0347)\}$
2	0.9312	LS and VS	$\{(D_1, 0.7682), (D_2, 0.2101), (D_3, 0), (D_4, 0), (D_5, 0.0119), (D_6, 0.0098)\}$
3	0.7482	LS and PS	$\{(D_1, 0.9902), (D_2, 0), (D_3, 0), (D_4, 0), (D_5, 0), (D_6, 0.0098)\}$
4	0.1305	LS and $M$	$\{(D_1, 1), (D_2, 0), (D_3, 0), (D_4, 0), (D_5, 0), (D_6, 0)\}$
14	0.1148	VS and $L$	$\{(D_1, 0.9753), (D_2, 0.0017), (D_3, 0), (D_4, 0), (D_5, 0), (D_6, 0.0230)\}$
15	0.2419	VS and ML	$\{(D_1, 0.8073), (D_2, 0.1317), (D_3, 0), (D_4, 0), (D_5, 0.0305), (D_6, 0.0305)\}$
16	0.3346	VS and VL	$\{(D_1, 0.9673), (D_2, 0.0019), (D_3, 0), (D_4, 0), (D_5, 0.0061), (D_6, 0.0247)\}$
17	0.1358	PS and LS	$\{(D_1, 0), (D_2, 0), (D_3, 0), (D_4, 0), (D_5, 0.2922), (D_6, 0.7078)\}$
52	0.3492	ML and $M$	$\{(D_1, 0.0016), (D_2, 0), (D_3, 0), (D_4, 0), (D_5, 0.1031), (D_6, 0.8953)\}$
53	0.6481	ML and PM	$\{(D_1, 0.0019), (D_2, 0.0019), (D_3, 0), (D_4, 0), (D_5, 0.1339), (D_6, 0.8625)\}$
54	0.3144	ML and $L$	$\{(D_1, 0), (D_2, 0), (D_3, 0), (D_4, 0), (D_5, 0), (D_6, 1)\}$
55	0.1654	ML and ML	$\{(D_1, 0.0019), (D_2, 0.0019), (D_3, 0), (D_4, 0), (D_5, 0.2589), (D_6, 0.7373)\}$
56	0.3144	ML and VL	$\{(D_1, 0.0019), (D_2, 0.0019), (D_3, 0), (D_4, 0), (D_5, 0.3422), (D_6, 0.6540)\}$
57	0.8946	VL and LS	$\{(D_1,0),(D_2,0),(D_3,0),(D_4,0),(D_5,0.0019),(D_6,0.9981)\}$
58	0.6879	VL and VS	$\{(D_1, 0), (D_2, 0), (D_3, 0), (D_4, 0), (D_5, 0), (D_6, 1)\}$
62	0.8467	VL and $L$	$\{(D_1, 0.0019), (D_2, 0.0019), (D_3, 0), (D_4, 0), (D_5, 0.2339), (D_6, 0.7623)\}$
63	0.5873	VL and ML	$\{(D_1, 0.0019), (D_2, 0.0019), (D_3, 0), (D_4, 0), (D_5, 0.4422), (D_6, 0.5540)\}$
64	0.9478	VL and VL	$\{(D_1, 0.0075), (D_2, 0.0059), (D_3, 0.0062), (D_4, 0.0055), (D_5, 0.5851), (D_6, 0.3898)\}$

Abbreviations: *L*, large; *LS*, limit small; *M*, medium; *ML*, medium large; *PM*, positive medium; *PS*, positive small; V-BRB, vector-based belief rule base; *VL*, very large; *VS*, very small.

*p*4, *p*6, and *p*8) are taken as the inputs of the BRB model (i.e., antecedent attributes), as shown in Figure 7. And the leak size is taken as the output of the BRB model (i.e., consequent attribute), as shown in Figure 6.

Thereby the V-BRB model can be constructed and optimized based on the proposed approach. Then, 510 samples are selected as the training data. It needs to be pointed out that these samples cover the data before, during, and after the leakage, which ensures the diversity of training data. Figures 8 and 9 show the inference effect of training data and all of the sample data, respectively.

It demonstrates that the estimated outcomes match the true values very closely. To further verify the superiority of the V-BRB, three classic BRB models using in the pipeline leak detection cases are selected for comparative analysis. The experimental results are shown in Table 6.

It can be seen from Table 6 that the V-BRB is better than the classic BRB models. First, the V-BRB is the most concise model since the indicator RN is the least. Correspondingly, the



FIGURE 4 Inference effect of training data after optimization [Color figure can be viewed at wileyonlinelibrary.com]



FIGURE 5 Inference effect of testing data after optimization [Color figure can be viewed at wileyonlinelibrary.com]

	Indicator							
			MSE					
Model	RN	OPN	Training data	Testing data	AIC			
O-BRB	256	1796	<u>0.1081</u>	0.1662	3624.91			
V-BRB	<u>64</u>	450	0.1142	0.1831	934.59			

ТАЕ	BLE	5	Experimental	results of	f O-BRB	and	V-BRB
-----	-----	---	--------------	------------	---------	-----	-------

*Note*: The bold underline means the value is better.

Abbreviations: AIC, Akaike Information Criterion; MSE, mean square error; O-BRB, original belief rule base; OPN, optimized parameter number; RN, rule number; V-BRB, vector-based belief rule base.



FIGURE 6 The curve of leak size [Color figure can be viewed at wileyonlinelibrary.com]



FIGURE 7 The inputs of the belief rule base [Color figure can be viewed at wileyonlinelibrary.com]

reduction of the number of rules reduces the parameters to be optimized, which can greatly improve the running speed of the model. Second, the MSE of the V-BRB is 0.3528 which means the accuracy of the V-BRB is the best. Third, by comparing AIC, it can be found that the V-BRB proposed in this paper improves the accuracy of the model while reducing the complexity of the model. In short, the V-BRB is more effective than other BRB models in dealing with the pipeline leakage detection data.

# 4.3 | Discussion

On the basis of the above two case studies and experimental results, the effectiveness and superiority of vector-based BRB approach were validated. It can be used to reduce the complexity of the BRB model without affecting its accuracy. For the inference process of vector-based BRB, the

5021





FIGURE 8 Inference effect of training data [Color figure can be viewed at wileyonlinelibrary.com]



FIGURE 9 Inference effect of the 2008 sample data [Color figure can be viewed at wileyonlinelibrary.com]

Models	Number of training data	RN	OPN	MSE (testing data)	AIC
Xu et al. <sup>18</sup>	500	56	336	0.4069	14137.06
Zhou et al. <sup>28</sup>	900	56	336	0.7880	15464.21
Chen et al. <sup>40</sup>	500	56	349	0.3990	14123.69
V-BRB	510	<u>32</u>	<u>226</u>	0.3528	13630.59

<b>IABLE 0</b> Experimental results of	of different	BKB	models
--	--------------	-----	--------

Note: The bold underline means the value is better.

Abbreviations: AIC, Akaike Information Criterion; BRB, belief rule base; MSE, mean square error; OPN, optimized parameter number; RN, rule number; V-BRB, vector-based belief rule base.

computational time can be ignored. During the modeling process of vector-based BRB, parameters optimization is necessary to achieve good performance. It takes up most of the computational time. Actually, in the same experimental environment, the optimization time mostly depends on the number of parameters. The more the number of parameters, the larger the optimization time. In case study 1, the optimized parameters are 1796 and 450, respectively. As a result, the optimization time of O-BRB and V-BRB is 53,212 and 7525 s, respectively. It indicates that the V-BRB is more than seven times faster than O-BRB. Similarly, in case study 2, our model outperforms other classical models. For example, the optimization time of the vector-based BRB model (i.e., 13,298 s) with 226 parameters is more than two times faster than that of Xu<sup>18</sup> (i.e., 27,054 s) with 336 parameters. In conclusion, our model is much faster in the same experimental environment.

# 5 | CONCLUSION

To overcome the shortcomings of the performance degradation caused by overnumbered antecedent attributes of the BRB model, a novel nonlinear causal inference approach based on vector-based BRB is proposed. The approach extends the numerical matching to multidimensional vector matching, which can effectively reduce the number of rules and thus streamline the scale of the BRB model. Meanwhile, the full activation method is used to calculate the similarity of the input attribute vector and all referential vectors. It not only improves the accuracy of the matching degree, but also retains the integrity and uncertainty of the information. The experimental results show that the proposed approach can greatly reduce the complexity of the model while ensuring the accuracy of the model.

Further experimental research can be conducted to optimize the V-BRB model. For example, the selection criteria of the model parameter *h* is worthy of study. In our experiments, we convert the antecedent attributes into two-dimensional (h = 2, four antecedent attributes in the case study 1) and three-dimensional (h = 3, six antecedent attributes in the case study 2) space vector, respectively. It is just one of the feasible strategies. Different values of *h* may affect the optimization process of the weight  $\delta_i$ . Therefore, it is a complicated problem which needs to be solved in the future work.

#### ACKNOWLEDGMENTS

We acknowledge financial support from the NSFC-Zhejiang Joint Fund for the Integration of Industrialization and Informatization, China (U1709215), Zhejiang Province Outstanding Youth Fund (R21F030005), Open Fund of National Engineering Research Centre for Water Transport Safety, China (A2020002), NSFC (61903108), Zhejiang Province Key R&D projects (2019C03104, 2018C04020, and 2021C03015), Zhejiang Province Public Welfare Technology Application Research Project (LGF20H270004 and LGF19H180018), Key project of Zhejiang Provincial Medical and Health Science and Technology Plan (WKJ-ZJ-2038).

#### ORCID

Zhang Zhenjie b https://orcid.org/0000-0001-7011-8889 Xu Xiaobin b https://orcid.org/0000-0003-1822-6190 Chen Peng b http://orcid.org/0000-0002-7473-4803 Wu Xudong b http://orcid.org/0000-0002-5578-7176 Xu Xiaojian b http://orcid.org/0000-0003-3179-1864 WILFY

# 5024 WILEY

Wang Guodong b http://orcid.org/0000-0003-0251-6257 Dustdar Schahram b http://orcid.org/0000-0001-6872-8821

#### REFERENCES

- 1. Yang JB, Liu J, Wang J, Sii HS, Wang HW. Belief rule-base inference methodology using the evidential reasoning approach—RIMER. *IEEE Trans Syst Man Cybern Part A Syst Hum.* 2006;36:266-285.
- 2. Yang JB, Xu DL. Nonlinear information aggregation via evidential reasoning in multiattribute decision analysis under uncertainty. *IEEE Trans Syst Man Cybern Part A Syst Hum.* 2002;32:376-393.
- 3. Dempster AP. A generalization of Bayesian inference. J Royal Statist Soc Ser B. 1968;30(2):205-247.
- 4. Shafer G. A Mathematical Theory of Evidence. Princeton, NJ: Princeton University Press; 1976.
- 5. Zadeh LZ. Fuzzy sets. Inf Control. 1965;8(3):338-353.
- 6. Pawlak Z. Rough sets. Int J Comput Inf Sci. 1982;11:341-356.
- 7. Deng Y. Uncertainty measure in evidence theory. Sci China-Inf Sci. 2020;63(11):210201.
- 8. Wen T, Deng Y. The vulnerability of communities in complex networks: An entropy approach. *Reliab Eng Syst Saf.* 2020;196:106782.
- 9. Zhang L, Zhan JM, Alcantud JCR. Novel classes of fuzzy soft-coverings-based fuzzy rough sets with applications to multi-criteria fuzzy group decision making. *Soft Comput.* 2019;23(14):5327-5351.
- 10. Song YF, Wang XD, Zhu JW, Lei L. Sensor dynamic reliability evaluation based on evidence theory and intuitionistic fuzzy sets. *Appl Intell.* 2018;48(11):3950-3962.
- 11. Liu ZG, Liu Y, Dezert J, Cuzzolin F. Evidence combination based on credal belief redistribution for pattern classification. *IEEE Trans Fuzzy Syst.* 2020;28(4):618-631.
- 12. Wen T, Duan SY, Jiang W. Node similarity measuring in complex networks with relative entropy. *Commun Nonlinear Sci Numer Simul.* 2019;78:104867.
- 13. Wen T, Deng Y. Identification of influencers in complex networks by local information dimensionality. *Inf Sci.* 2020;512:549-562.
- 14. Wen T, Pelusi D, Deng Y. Vital spreaders identification in complex networks with multi-local dimension. *Knowl-Based Syst.* 2020;195:105717.
- 15. Sun R. Robust reasoning: Integrating rule-based and similarity-based reasoning. *Artif Intell.* 1995;75: 241-295.
- 16. Liu J, Yang JB, Ruan D, Martinez L, Wang J. Self-tuning of fuzzy belief rule bases for engineering system safety analysis. *Ann Oper Res.* 2008;163(1):143-168.
- 17. Jiang J, Li X, Zhou ZJ, Xu DL, Chen YW. Weapon system capability assessment under uncertainty based on the evidential reasoning approach. *Expert Syst Appl.* 2011;38(11):13773-13784.
- 18. Xu DL, Liu J, Yang JB, et al. Inference and learning methodology of belief-rule-based expert system for pipeline leak detection. *Expert Syst Appl.* 2007;32:103-113.
- 19. Yang JB, Wang YM, Xu DL, Chin KS, Chatton L. Belief rule-based methodology for mapping consumer preferences and setting product targets. *Expert Syst Appl.* 2012;39:4749-4759.
- 20. Zhou ZJ, Hu CH, Han XX, He HF, Ling XD, Zhang BC. A model for online failure prognosis subject to two failure modes based on belief rule base and semi-quantitative information. *Knowl-Based Syst.* 2014;70:221-230.
- 21. Kong GL, Xu DL, Body R, Yang JB, Kevin MJ, Carley S. A belief rule-based decision support system for clinical risk assessment of cardiac chest pain. *Eur J Oper Res.* 2012;219:564-573.
- 22. Sachan S, Yang JB, Xu DL, Benavides DE, Li Y. An explainable AI decision-support-system to automate loan underwriting. *Expert Syst Appl.* 2019;144:113100.
- 23. Hossain MS, Rahaman S, Mustafa R, Andersson K. A belief rule-based expert system to assess suspicion of acute coronary syndrome (ACS) under uncertainty. *Soft Comput.* 2018;22:7571-7586.
- 24. Calzada A, Liu J, Wang H, Nugent C, Martinez L. Application of a spatial intelligent decision system on self-rated health status estimation. *J Med Syst.* 2015;39(11):1-18.
- 25. Abudahab K, Xu DL, Chen YW. A new belief rule base knowledge representation scheme and inference methodology using the evidential reasoning rule for evidence combination. *Expert Syst Appl.* 2016;51: 218-230.
- 26. Yang JB. Rule and utility based evidential reasoning approach for multiple attribute decision analysis under uncertainty. *Eur J Oper Res.* 2001;131(1):31-61.

- 27. Chang LL, Zhou Y, Jiang J, Li MJ, Zhang XH. Structure learning for belief rule base expert system: A comparative study. *Knowl-Based Syst.* 2013;39:159-172.
- Zhou ZJ, Hu CH, Yang JB, Xu DL, Chen MY, Zhou DH. A sequential learning algorithm for online constructing belief-rule-based systems. *Expert Syst Appl.* 2010;37(2):1790-1799.
- 29. Wang YM, Yang LH, Chang LL, Fu YG. Rough set method for rule reduction in belief rule base. *Control Decis.* 2014;29(11):1943-1950.
- 30. Yang LH, Ye FF, Wang YM. Ensemble belief rule base modeling with diverse attribute selection and cautious conjunctive rule for classification problems. *Expert Syst Appl.* 2020;146:113161.
- Li GL, Zhou ZJ, Hu CH, Chang LL, Zhou ZG, Zhao FJ. A new safety assessment model for complex system based on the conditional generalized minimum variance and the belief rule base. *Safety Sci.* 2017;93: 108-120.
- 32. Chang LL, Zhou ZJ, Chen YW, Xu XB, Sun JB, Liao TJ. Akaike information criterion-based conjunctive belief rule base learning for complex system modeling. *Knowl-Based Syst.* 2018;161:47-64.
- Savan EE, Yang JB, Xu DL, Chen YW. A genetic algorithm search heuristic for belief rule-based modelstructure validation. In: *Proceedings of the International Conference on Systems, Man, and Cybernetics*. Manchester, UK: IEEE; 2013:1373-1378.
- 34. Akaike H. Akaike's Information Criterion. Heidelberg, Berlin: Springer; 2011.
- Chang LL, Chen YW, Hao ZY, Zhou ZJ, Xu XB, Tan X. Indirect disjunctive belief rule base modeling using limited conjunctive rules: Two possible means. *Int J Approx Reasoning*. 2019;108:1-20.
- Chang LL, Zhou ZJ, You Y, Yang LH, Zhou ZG. Belief rule based expert system for classification problems with new rule activation and weight calculation procedures. *Inf Sci.* 2016;336:75-91.
- 37. Yang LH, Wang YM, Chang LL, Fu YG. A disjunctive belief rule-based expert system for bridge risk assessment with dynamic parameter optimization model. *Comput Ind Eng.* 2017;113:459-474.
- Chang LL, Zhou ZJ, Liao HC, Chen YW, Tan X, Herrera F. Generic disjunctive belief rule base modeling, inferencing, and optimization. *IEEE Trans Fuzzy Syst.* 2019;27(9):1866-1880.
- 39. Abdi H, Williams LJ. Principal component analysis. Wiley Interdiscip Rev: Comput Stat. 2010;2(4):433-459.
- Chen YW, Yang JB, Xu DL, Zhou ZJ, Tang DW. Inference analysis and adaptive training for belief rule based systems. *Expert Syst Appl.* 2011;38(10):12845-12860.

**How to cite this article:** Zhenjie Z, Xiaobin X, Peng C, et al. A novel nonlinear causal inference approach using vector-based belief rule base. *Int J Intell Syst.* 2021;36: 5005-5027. https://doi.org/10.1002/int.22500

# APPENDIX: THE LIST OF NOTATIONS

Notation	Introduction
Rule <sub>k</sub>	kth rule
$x_i$	Antecedent attribute
$T_k$	Number of the antecedent attributes
$\theta_k$	Rule weight
RN	Number of rules
$\delta_i$	Weight of the <i>i</i> th antecedent attribute
$A_i^k$	Referential value of the <i>i</i> th antecedent attribute in the <i>k</i> th rule
$A_{i,j}$	<i>j</i> th referential value of the <i>i</i> th antecedent attribute
$v_i$	Number of referential values of the <i>i</i> th antecedent attribute
$D_n$	Referential value of the consequent attribute
Ν	Number of referential values of the consequent attribute
$\omega_k$	Activation weight
$\overline{\delta_i}$	Normalization of antecedent attribute weights
$\alpha_i^k$	Matching degree of the referential value of the $i$ th antecedent attribute in the $k$ th rule
$lpha_{i,j}^k$	Matching degree of the $j$ th referential value of the $i$ th antecedent attribute in the $k$ th rule
$\beta_n^k$	Initial belief degree of corresponding referential level determined by expert knowledge
$\hat{\beta}_n$	Belief degree of the output referential value of $D_n$ after fusion
μ	Intermediate parameters
У	Output of BRB
X	Input data set (input matrix)
Т	Sampling values of the <i>r</i> th attribute variable
J	Number of attribute subvectors
$F_i$	<i>i</i> th principal component
w <sub>ri</sub>	Weight coefficient of the <i>r</i> th attribute for the <i>i</i> th principal component
W	Transformation matrix
$\tau(r)$	Mean value of attribute $x_r$
$\tilde{X}$	Input matrix after centralization
V	Correlation matrix
$\lambda_r$	rth eigenvalue
$\varphi_r$	Contribution rate of the <i>r</i> th principal component $F_r$
h	Number of original attributes contained in each vector
$K_i$	Number of cluster centers corresponding to the <i>i</i> th group of data
$n(C_{k,i})$	Number of elements in the cluster $C_{k,i}$

(Continues)

Notation	Introduction
$A_r$	rth referential vector
$R_r$	rth input vector
$d_r^{K_r}$	Distance of $R_r$ and $A_r$
с	Number of original attributes in the last attribute vector
$g_r^{K_r}$	Reciprocal of $d_r^{K_r}$
$\alpha_r^{K_r}$	Normalization of $g_r^{K_r}$
0	Observation value of the actual sample set
Ô	Actual output value of the BRB model
Р	Parameter set to be optimized
<i>ξ</i> ( <i>P</i> )	Error between $O$ and $\hat{O}$
$M_p$	Number of the referential vectors
$C_{RN}$	Rule reduction coefficient
OPN	Number of optimized parameters
MSE	Mean square error between estimated values and actual values
AIC	Akaike Information Criterion