Knowledge-Based Systems 240 (2022) 107805

Contents lists available at ScienceDirect

Knowledge-Based Systems

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

Highly explainable cumulative belief rule-based system with effective rule-base modeling and inference scheme

Long-Hao Yang [a,c], Jun Liu [c], Fei-Fei Ye [b,c,*], Ying-Ming Wang [a], Chris Nugent [c], Hui Wang [c], Luis Martínez [d]

[a] Decision Sciences Institute, Fuzhou University, Fuzhou 350108, China
[b] School of Public Administration, Fujian Normal University, Fuzhou, 350117, China
[c] School of Computing, Ulster University at Jordanstown Campus, Newtownabbey BT37 0QB, Northern Ireland, UK
[d] Department of Computer Science, University of Jaén, Spain

ARTICLE INFO

Article history:
Received 30 July 2021
Received in revised form 25 October 2021
Accepted 20 November 2021
Available online 20 December 2021

Keywords:
Rule-based system
Uncertainty
Belief rule base
Knowledge representation
Information fusion
Explainable AI

ABSTRACT

Advancement and application of rule-based expert systems have been a key research area in explainable artificial intelligence (XAI) because the rule-base is one of the most common and natural explainable frameworks for knowledge representation. The present work aims to design a novel rule-based system, called Cumulative Belief Rule-Based System (CBRBS), by establishing efficient rule-base modeling and inference procedures, where the rule-base modeling procedure includes the generation of cumulative belief rules via numeric data transformation and extended belief rule integration, and the rule-base inference procedure includes the inference of cumulative belief rules via consistent rule activation and activated rule integration. All these procedures enable CBRBS to better achieve the balance of explainability, high-efficiency (or computing complexity), and accuracy to fit with different application scenarios, as well as overcome the limitations of classical rule-based systems. Extensive experiments based on the well-known pipeline leak detection problem and open-source classification problems are conducted to illustrate the feature and advantage of the CBRBS over other classical rule-based systems and some commonly used classifiers.

1. Introduction

The recent rapid progress in machine learning has led to resurgence in interest in explainable artificial intelligence (XAI). Hagras [1] summarized three approaches to realize XAI: (1) deep explanation: the use of deep learning techniques to learn explainable features, like some successful achievements in automatic ship classification [2] and pattern recognition applications [3]; (2) interpretable models: techniques to learn more structured, interpretable, and causal models, *e.g.*, rule-based modeling whose latest achievements include interpretable fuzzy classifier for big data [4] and deep convolutional fuzzy systems (DCFS) to stock index prediction [5]; and, (3) model induction: techniques to infer explainable model from any model as a black box, *e.g.* critical thinking about XAI for rule-based fuzzy systems [6]. This paper focuses on the second approach for an effective rule-based system.

Providing an effective rule-based system is a general trend in XAI and has attracted a lot of attention and researches from different fields in order to support human in judging the quality of the assessment in the decision process, because rule-based systems represent knowledge in terms of rules that tell decision-makers what you should do or what you could conclude in different situations [7]. Compared to black box approaches that use no self-explanatory modeling, rule-based systems foster understanding the underlying problem and the reasons behind intelligent decisions and predictions, *i.e.*, its explanation capability [8]. To make a black-box scheme into a white box is rather challenging. An alternative and feasible route in term of XAI is to, starting from a white-box system, try to enhance significantly their performance in terms of accuracy and efficiency (or computing complexity) without loss of transparency, interpretation, and explainability. This forms one of the key motivations of the proposed work.

The core of a rule-based system is typically a set of IF-THEN rules, which contains the knowledge provided by domain experts and/or extracted from historical data, and an inference engine, which allows the system to produce an output using IF-THEN rules. With the emergence of white-box modeling of XAI techniques, a successful rule-based system is required to have a great potential for obtaining a favorable trade-off between accuracy, efficiency, and explainability. Although a lot of off-the-shelf rule-based systems were widely used in different areas [9,10], their weaknesses such as opaque relations between

* Corresponding author.
E-mail address: 946755430@qq.com (F.-F. Ye).

https://doi.org/10.1016/j.knosys.2021.107805
0950-7051/

rules, inference efficiency problems, ineffective search strategy, and knowledge acquisition bottleneck have limited their market value and still some challenging issues [11]. This work aims to make good contributions by advancing the existing rule-based systems and overcoming their limitations without losing their original explainability capability.

In the past development of rule-based systems, the system based on the fuzzy logic, including fuzzy sets, fuzzification, and defuzzification, to generate IF-THEN rules and implement the inference engine firstly gained attentions [12]. Such a system is so called a fuzzy rule-based system (FRBS). From the first FRBS proposed by Mamdani and Assilian [13] for a model industrial plant, Mamdani-type FRBS was regarded as an explainable system modeling tool and has been studied in a wide variety of domains [14–17]. Within the Mamdani-type FRBS framework, it is important to note that Wang and Mendel [18] proposed a non-iterative method to generate fuzzy rules by learning from data and this kind of FRBS (WM-FRBS for short) is a milestone in the development of Mamdani-type FRBSs and data-driven model, because it achieves the balance between accuracy, efficiency, and explainability to a new level.

In 2006, in order to enhance the knowledge representation of IF-THEN rules, a belief structure was embedded into the THEN part, which is so called belief rule [19]. Comparing with a fuzzy rule, a belief rule can provide a better explainability because of its capability in representing fuzzy, random, and incomplete uncertainties [20]. Moreover, the evidential reasoning (ER) approach is used as an inference engine to develop a belief rule-based system (BRBS). However, the combinatorial explosion problem was a quite frequent keyword and is a great obstacle in BRBS [21,22]. This has been a common challenge for other traditional rule-based systems including Mamdani-type FRBS. For overcoming the combinatorial explosion problem, an extended BRBS (EBRBS) [23] was developed by embedding belief structures into all antecedent attributes of the belief rule (so called extended belief rule) and also proposing an efficient method to generate extended belief rule base (EBRB). The main feature of EBRBS is that EBRB imports all benefits from BRB and also extends them to achieve new advantages. However, an EBRBS must address the challenges that have been mentioned in many previous studies: (1) the one-to-one correspondence between rules and data would lead to the rule boundlessness of the EBRB when facing continuous supply of data; (2) the EBRB inference always suffers from the rule inconsistency problem that all extended belief rules in an EBRB are almost activated to produce an inference output for any given input data.

To address the mentioned challenges, many endeavors have been undertaken in the last few years. The detailed review of the endeavors can be found in Section 2.3. There is no doubt that all previous endeavors have contributed enormously to the development of EBRBS and, comparatively, Micro-EBRBS has already advantages over EBRBS in terms of accuracy, efficiency, and explainability; especially it is not required to include a time-consuming iterative process when solving the challenges. Hence, on the basis of Micro-EBRBS, an advanced rule-based system, called *Cumulative Belief Rule-Based System* (CBRBS), is proposed in the present study and also deal with the questions needed to be solved in order to improve Micro-EBRBS, *e.g.*, How can a Micro-EBRBS effectively activate rules in its inference process? What is mathematical basis for a Micro-EBRBS? How can a Micro-EBRBS be efficiently generated from input–output data pairs? For these reasons, the focus of the present study aims to establish a generic rule-base modeling and inference procedures for CBRBS, set up the generic methodological framework with its key features and theoretical foundations established.

More specifically, as the first procedure in the rule-base modeling and its inference, the transformation of inputs into belief distributions is discussed to propose a new function in the case of numeric data, which provides a new perspective for input transformation according to the distance relationship between numeric data and utility values. Based on the efficient rule generation method used in EBRBS, the clustering and fusion of extended belief rules to generate cumulative belief rules using the ER algorithm are proposed to develop a cumulative BRB (CBRB) modeling procedure. Additionally, by splitting the determination of activated rules and the calculation of activation weights in the EBRB inference procedure, a nearest neighbor strategy is proposed on the basis of belief distributions to activate consistent cumulative belief rules firstly, followed by the integration of activated rules using the ER algorithm to produce inference outputs. The integrated process of rule activation forms a CBRB inference procedure. Hence, compared to Micro-EBRBS [24], CBRBS has the ability to not only solve the rule boundlessness of EBRB modeling, but also address the rule inconsistency of EBRB inference.

The main contributions of the present work can be summarized as follows:

(1) An effective rule-base modeling procedure is proposed to efficiently generate CBRB from data samples. The most important thing is that this procedure can overcome the challenge of rule boundlessness and combinatorial explosion.

(2) An effective rule-base inference procedure is proposed to efficiently produce an output based on CBRB. The new procedure is able to overcome the challenge of rule inconsistency without time-consuming rule traversal process.

(3) The common inherent properties, including linear computing complexity, boundedness, and continuity properties, and universal approximation theorem are investigated for the first time to facilitate the application of CBRBS.

(4) One well-known regression problem, named oil pipeline leak detection, and eighteen open-source classification problems are used together to provide abundant experimental analyses to demonstrate the effectiveness of CBRBS.

The rest of the present study is organized as follows: Section 2 discusses the background of EBRBS and typical rule-based systems. Section 3 and Section 4 propose efficient modeling and inference procedures to upgrade an EBRBS as a CBRBS. Section 5 illustrates the framework and properties of CBRBS. Section 6 provides comparative experiments to demonstrate the advantage of CBRBS. And the paper is concluded in Section 7.

2. Background of EBRBS and typical rule-based systems

In this section, the discussion of belief structure and rule representation is firstly provided, followed by the introduction of EBRBS and its related works. Then typical rule-based systems are introduced in the purpose of comparisons.

2.1. Rule representation based on belief structure and its explainability

A belief structure is a distributed assessment with belief degrees. The origin of the belief structure comes from the Dempster–Shafer theory [25,26], which is a powerful and flexible mathematical tool for addressing imprecise and uncertain information, so the belief structure is also known as a basic belief assignment under a collectively exhaustive and mutually exclusive set of propositions. Thereafter, the belief structure was used in the ER approach [27] for multi-attribute decision analysis (MADA) to model a subjective assessment with uncertainty.

Suppose N distinctive (mutually exclusive) evaluation grades $\{D_n;\ n = 1,\ldots, N\}$, *e.g.*, {*Zero*, *Middle*, *Large*}, are used to assess attribute D of a certain problem, *e.g.*, leak size in an pipeline leak detection problem [28]. Based on actual data and expert experience, a belief structure and its corresponding belief distribution can be denoted as follows:

$$S(D) = \{(D_n, \beta_n);\ n = 1, \ldots, N\} => S(Leak\,size) = \{(Zero, 0.3), (Middle, 0.6), (Large, 0)\} \quad (1)$$

where β_n denotes the belief degree assigned to evaluation grade D_n, and it satisfies $0 \leq \beta_n \leq 1$ and $\sum_{n=1}^{N} \beta_n \leq 1$. Specifically, the leak size of the oil pipeline leak detection problem is assessed to be *Zero* with belief degree 0.3, *Middle* with belief degree 0.6, and *Large* with belief degree 0. More importantly, due to $0.3 + 0.6 + 0 = 0.9 < 1.0$, the remaining belief degree $1.0 - 0.9 = 0.1$ can be regarded as global uncertain information and kept in the belief structure.

Owing to the belief structure, an IF-THEN rule has three types of representation schemes as follows:

(1) The first one is the rule representation without belief structures, which is the popular scheme used in Mamdani-type FRBSs, *e.g.*, WM-FRBS [18] and fuzzy rule-based classification system (FRBCS) [29]. Assume J_i fuzzy sets (evaluation grades) $\{A_{i,j};\ j = 1, \ldots, J_i\}$ are used to describe the ith ($i = 1, \ldots, M$) antecedent attribute U_i as well as N fuzzy sets $\{D_n;\ n = 1, \ldots, N\}$ for consequent attribute D. Therefore, a fuzzy rule can be written as follows:

$$R_k: IF\ \ U_1 is\ A_1^k \wedge \cdots \wedge U_M\ \ is\, A_M^k,\ THEN\ \ D\, is\, D_k,\ \ with\, \theta_k \quad (2)$$

where $A_i^k \in \{A_{i,j};\ j = 1, \ldots, J_i\}$ and $D_k \in \{D_n;\ n = 1, \ldots, N\}$; θ_k denotes the weight of the kth fuzzy rule.

(2) The second one is the rule representation with the belief structure used in consequent attribute, which provides an informative and realistic scheme for rule representation and is referred to as a belief rule [19]. Comparing with fuzzy rules, the evaluation grades used to describe the M antecedent attributes and one consequent attribute are called referential values and consequents, respectively. Consequently, a belief rule can be written as follows:

$$R_k: IF\ \ U_1 is\, A_1^k \wedge \cdots \wedge U_M\ is\, A_M^k,\ THEN\ \ D\ \ is\ \ \{(D_n, \beta_n^k);\ n{=}1, \ldots, N\},\ with\ \theta_k \quad (3)$$

where $\{(D_n, \beta_n^k);\ n = 1, \ldots, N\}$ is the belief structure in the kth belief rule, β_n^k is the belief degree assigned to consequent D_n.

(3) The third one is the rule representation with the belief structure used in antecedent and consequent attributes, which further extends the ability of fuzzy rules to provide explainable information representation in their rule antecedent and is so called an extended belief rule [23]. Based on the given referential values $\{A_{i,j};\ j = 1, \ldots, J_i\}$ and consequents $\{D_n;\ n = 1, \ldots, N\}$, an extended belief rule can be written as follows:

$$\begin{aligned} R_k: &IF\ U_1\ \ is\ \ \{(A_{1,j}, \alpha_{1,j}^k);\ j = 1, \ldots, J_1\} \wedge \cdots \wedge U_M\ \ is\ \ \{(A_{M,j}, \alpha_{M,j}^k);\\ &j = 1, \ldots, J_M\},\\ &THEN\ D\ is \{(D_n, \beta_n^k);\ n = 1, \ldots, N\},\ with\, \theta_k \end{aligned} \quad (4)$$

where $\{(A_{i,j}, \alpha_{i,j}^k);\ j = 1, \ldots, J_i\}$ is the belief structure used in the ith antecedent attribute at the kth extended belief rule, $\alpha_{i,j}^k$ is the belief degree assigned to referential value $A_{i,j}$.

From the above three IF-THEN rules, a fuzzy rule is a special case of a belief rule and the latter one is another special case of an extended belief rule in the term of rule representation. For example, when the belief distribution in M antecedent attributes satisfies $\alpha_{i,j}^k = 1$ ($j \in \{1, \ldots, J_i\}$; $i = 1, \ldots, M$) and $\alpha_{i,t}^k = 0$ ($t = 1, \ldots, J_i$; $t \neq j$), the extended belief rule can be simplified into the belief rule. Similarly, when the belief distribution in consequent attribute satisfies $\beta_n^k = 1$ ($n \in \{1, \ldots, N\}$) and $\beta_t^k = 1$ ($t = 1, \ldots, N$; $t \neq n$), the belief rule can be simplified into the fuzzy rule. Additionally, the extended belief rule has more benefits for overcoming the curse of dimensionality because each rule contains all evaluation grades of all attributes, which means that it is unnecessary to traverse all evaluation grades of all attributes for constructing a complete rule-base. This has been a common challenge, named combinatorial explosion, for other traditional rule-based systems in the form of fuzzy rule and belief rule. For example, there are J evaluation grades for M attributes, resulting in a total of J^M fuzzy rules or belief rules when a complete rule-base is required to construct. It is obvious that the size of the rule-base would grow exponentially along with the increase of the attributes. More details of example analyses regarding combinatorial explosion problem in the form of different rule representations can be found in Section 3.3.

2.2. Brief introduction to EBRBS and its challenges

Through taking into account extended belief rules in a rule-base, an EBRB can be obtained and hereby acts as the knowledge base of an EBRBS. According to [23], the generation of an EBRB is not based on any optimization technique, namely, each extended belief rule is generated from an input–output data pair using a non-iteration method. Hence, given a set of input–output data pairs of the form $< \boldsymbol{x_k}, y_k> = < x_{k,1}, \ldots, x_{k,M}, y_k>$ ($k = 1, \ldots, T$) and new input data of the form $\boldsymbol{x} = (x_1, \ldots, x_M)$, the EBRB modeling and the EBRB inference are shown in Fig. 1 and they include the following three steps:

Step 1: Transformation of inputs into belief distributions. When referential values $\{A_{i,j};\ j = 1, \ldots, J_i\}$ and consequents $\{D_n;\ n = 1, \ldots, N\}$ are provided for given antecedent attributes $\{U_i;\ i = 1, \ldots, M\}$ and consequent attribute D, the corresponding input data $x_{k,i}$ and output data y_k can be transformed into belief distributions $S(U_i, x_{k,i}) = \{(A_{i,j}, \alpha_{i,j}^k);\ j = 1, \ldots, J_i\}$ and $S(D, y_k) = \{(D_n, \beta_n^k);\ n = 1, \ldots, N\}$ using the definitions found in Section 3.1.

Step 2: Calculation of rule weights based on belief distributions. After obtaining T groups of belief distributions $<S(U_1, x_{k,1}), \ldots, S(U_M, x_{k,M}), S(D, y_k)>$ ($k = 1, \ldots, T$), the kth group of belief distributions forms a basic framework of the kth extended belief rule, namely, *IF* U_1 is $S(U_1, x_{k,1}) \wedge \cdots \wedge U_M$ is $S(U_M, x_{k,M})$, *THEN* D is $S(D, y_k)$. Consequently, a rule weight θ_k can be calculated based on the similarity of belief distributions in rule antecedent and rule consequent as follows:

$$\theta_k = 1 - \frac{Incons(R_k)}{\sum_{j=1}^{T} Incons(R_j)} \quad (5)$$

where $Incons(R_k)$ denotes the inconsistency degree of the kth extended belief rule and it is calculated by the similarity of belief distributions in rule antecedent and rule consequent as follows:

$$Incons(R_k) = \sum_{l=1, l \neq k}^{T} \left(1 - \exp\left\{ - \frac{\left(\frac{\min_{i=1,\ldots,M}\{Sim(U_i, x_{l,i}, x_{k,i})\}}{Sim(D, y_l, y_k)} - 1 \right)^2}{\left(\frac{1}{\min_{i=1,\ldots,M}\{Sim(U_i, x_{l,i}, x_{k,i})\}} \right)^2} \right\} \right) \quad (6)$$

where $Sim(\cdot, \cdot, \cdot)$ denotes the similarity of two belief distributions and its definition can be found in Section 4.2.

Step 3: Activation of extended belief rules to reply new input data. For replying new input data $\boldsymbol{x} = (x_i;\ i = 1, \ldots, M)$, any extended belief rule in the EBRB may be activated through (1) the transformation of new input data into belief distributions, (2) the similarity calculation of belief distributions, and (3) the calculation of activation weights for extended belief rules, namely w_k. Finally, the belief distribution $\{(D_n, \beta_n^k);\ n = 1, \ldots, N\}$ of

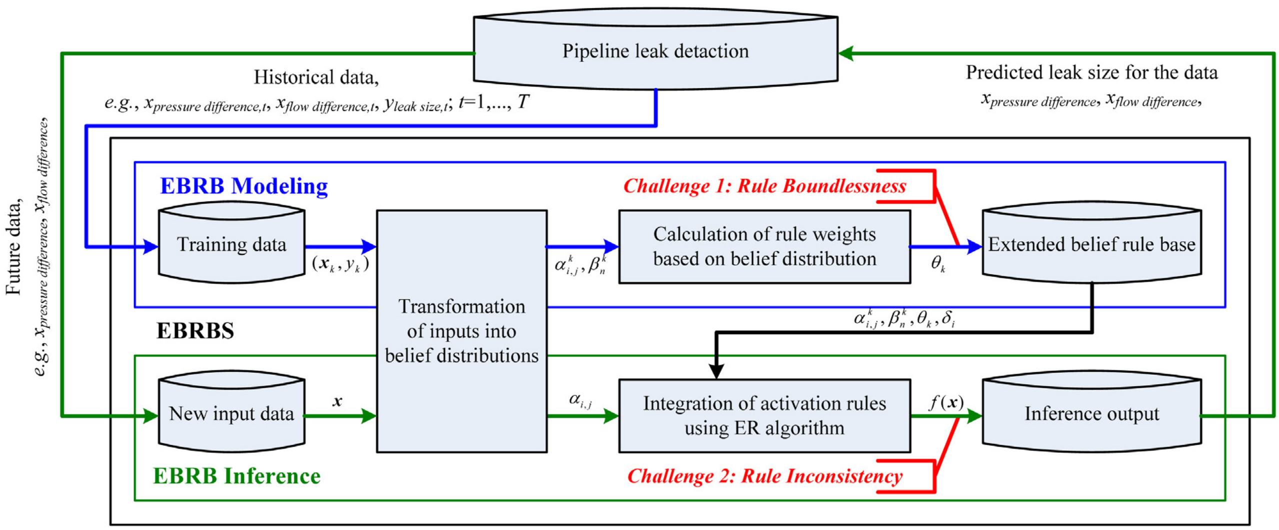


Fig. 1. Framework of EBRBS and its modeling and inference processes.

$$ER_n(\beta^k_s, w_k) = \frac{\prod^T_k(w_k\beta^k_n + 1 - w_k\sum^N_{s=1}\beta^k_s) - \prod^T_k(1 - w_k\sum^N_{s=1}\beta^k_s)}{\sum^N_{t=1}\prod^T_k(w_k\beta^k_t + 1 - w_k\sum^N_{s=1}\beta^k_s) - (N-1)\prod^T_k(1 - w_k\sum^N_{s=1}\beta^k_s) - \prod^T_k(1 - w_k)} \tag{7}$$

Box I.

all rules should be aggregated to produce an inference output, *e.g.*, the nth aggregated belief degree β_n via the analytical ER algorithm is expressed as given in Box I .

From the above introduction of the EBRBS, two challenges are summarized below needed to be overcome.

Challenge 1 (*Rule boundlessness of EBRB Modeling, [30]*)**.** When a large number of training data are provided and should be used for constructing an EBRB, the size of the EBRB must be overlarge because the EBRB modeling satisfies the one-to-one correspondence between extended belief rules and input–output data pairs.

Challenge 2 (*Rule Inconsistency of EBRB Inference, [31]*)**.** All extended belief rules in an EBRB have to be activated for producing inference outputs because the similarity of belief distributions between input data and rule antecedent is always greater than zero. Consequently, the EBRB inference is based on the conflicting information to reply new input data.

In order to explain Challenges 1 and 2, Fig. 2 provides a data analysis based on an existing case study of the oil pipeline leak detection problem [28] which totally has 2008 sample data. From Fig. 2, when different numbers of training data selected from all sample data are required to generate extended belief rules and activate these rules for replying the 2008 sample data, the corresponding numbers of extended belief rules and activated rules can clearly demonstrate that (1) the size of an EBRB is equal to the number of training data; (2) almost all extended belief rules have to be activated to reply any input data. Both of them are remarkable features of Challenges 1 and 2.

2.3. Related works on addressing the challenges of EBRBS

In order to address and overcome the two challenges pointed out in Section 2.2, many endeavors have been undertaken in recent years. These can be categorized into the following two types:

(1) Rule generation adjustment for optimizing EBRB construction, which aims to generate useful extended belief rules from historical input–output data pairs when constructing an EBRB. In this respect, the representative studies include: Jin et al. [32] designed a new parameter called certainty degrees and considered the parameter in the construction of EBRB. They concluded that the new EBRB offered a further improvement and a great extension of the original one. Yang et al. [30] utilized the activation weights and the contribution degrees of each extended belief rule to construct a novel decision-making-unit, followed by a data envelopment analysis (DEA) model to evaluate the effectiveness of each extended belief rule. Results showed that the proposed method can downsize EBRB and improve the accuracy of EBRBS. Similar to the use of DEA model, Ye et al. [33] applied an interval DEA model to calculate the reliability of each extended belief rule so that the construction of an EBRB could take into consideration the quality of data source. The real case study demonstrated that the resulting EBRBS was able to distinguish regional differences in the efficiency of environmental pollution management. Zhang et al. [34] proposed an EBRB reduction and parameter training method based on the DBSCAN algorithm, which has the function of integrating redundant rules and selecting effective data to generate rules. They suggested that the proposed method has a good performance on reducing EBRB's scale and improving EBRBS's efficiency. A similar study to propose a parameter training method for optimizing EBRB construction could be found in [35]. Afterwards, Espinilla et al. [36] used dependence measure and consistency measure to perform structure training for constructing an effective EBRB. The case study of activity recognition showed that the structure training was an important and necessary process in optimizing EBRB construction. By integrating the EBRB's parameter training and structure training, Wang et al. [37] proposed a joint training method to optimize the number and

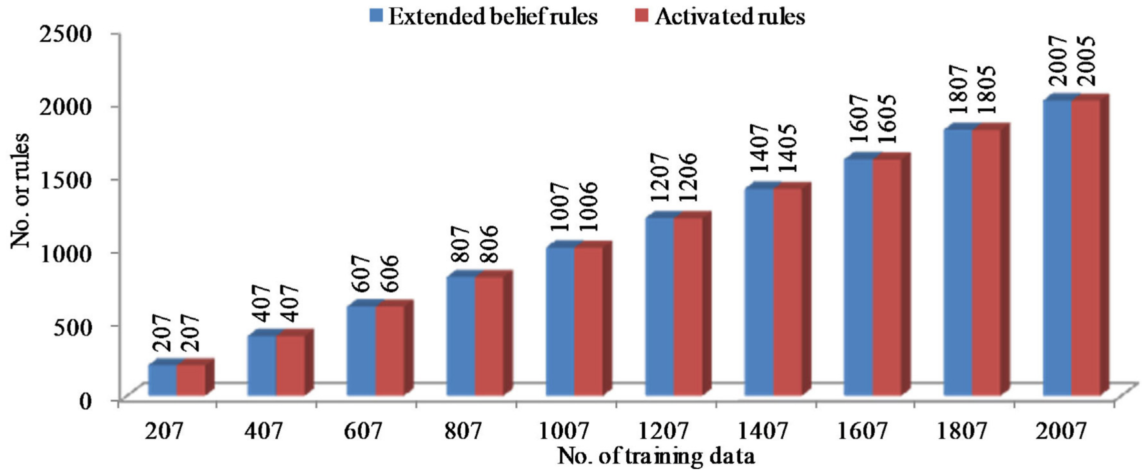


Fig. 2. Comparison of numbers of training data, extended belief rules and activated rules.

the value of parameters when constructing an EBRB for environmental governance cost prediction. Recently, Yang *et al.* [24] presented a domain division-based rule reduction method to combine extended belief rules using the similar belief distribution of rule antecedent and the reduced EBRBS is therefore named as Micro-EBRBS. They demonstrated that the Micro-EBRBS has excellent accuracy and efficiency on some big data classification problems.

(2) Rule activation adjustment for optimizing EBRB inference, which aims to select representative activated extended belief rules, instead of almost all rules, to reply any given new input data. In this respect, Calzada et al. [31] firstly pointed out the issues of incompleteness and inconsistency in EBRBS because of inherent defects in the process of rule activation, so a dynamic rule activation (DRA) method was proposed to activate extended belief rules in an iterative way. Afterwards, Espinilla et al. [36] further applied the DEA-based rule activation method as a core technique to enhance the performance of EBRBS-based activity recognition model. The results showed that the adaptation of the EBRBS-based model for smart environment provided an encouraged performance against the most popular classifiers. Yang et al. [38] extended the DRA method in term of determining the best set of activated rules, and two tree structures were introduced to develop a multi-attribute search framework (MaSF) for better rule activation. Following the tree structure to optimizing rule activation, several similar studies, such as the VP-tree-based and MVP-tree-based rule activation methods propose by Lin et al. [39], the bass tree-based rule activation method proposed by Fu et al. [40], were carried out and the corresponding experimental studies all demonstrated that these tree structure-based rule activation methods not only could increase the classification accuracy of EBRBS, but also have the advantages in increasing the efficiency of EBRB inference. Subsequently, on the basis of DEA models to evaluate the relative effectiveness of each activated extended belief rules, a consistency analysis-based method [41] was proposed to select representative activated rules, respectively, and the comparative studies showed that the DEA-based rule activation methods could improve the accuracy and rule activation rate of EBRBS to different degrees. Recently, Zhu et al. [42] aimed at solving high computational costs and long response times in EBRBS's rule activation process. Thus they proposed a minimum center distance rule activation (MCDRA) method to efficiently and accurately activate extended belief rules. Furthermore, an enhanced MCDRA method was proposed in [43] to demonstrate that the MCDRA method was able to obtain satisfactory rule activation ratios, accuracies, and response time compared with other rule activation methods. Fang et al. [44] focused on the use of EBRBS to handle imbalanced classification problems. They proposed a balance adjusting approach to adjust the activation weights of all activated rule for balancing the sums of activation weights belonging to the majority and the minority classes.

To summarize, the aforementioned endeavors solved the two challenges of EBRBS to different degrees. However, the existing studies on rule activation adjustment may be limited by computing efficiency or distance measure of the proposed methods. Comparatively, the existing studies on rule generation adjustment lack enough attentions but it should not be neglected because the big data is a common challenge to system modeling. Only by constructing Micro-EBRBS can address the challenge of rule boundlessness. Hence, the main work of the present study focuses on proposing an advanced rule-based system with efficient modeling and inference procedures based on the Micro-EBRBS.

2.4. Other typical rule-based systems for comparisons

The WM-FRBS [18] is a milestone in the development of Mamdani-type FRBSs and data-driven models, whose rule representation is based on the fuzzy rule shown in Eq. (2). Over time, the efficiency of WM-FRBS was successfully proven in many real-world applications [17,29,45], and its basic idea of using historical data only one-pass to generate a FRBS is considered as an important modeling framework in the fuzzy community. Therefore, the WM-FRBS is regarded as one of benchmarks to compare with our study. Basically, the details of WM-FRBS can be found in [18].

The BRBS [19] is a new rule-based system with the characteristic of both expert systems and data-driven models, whose rule representation is based on the belief rule shown in Eq. (3). Owing to the ability of belief rules to express various types of uncertainties, including fuzziness, randomness, and ignorance, BRBS has been applied in numerous fields [46,47], and many researchers were devoted to the improvement of BRBS in the terms of accuracy, interpretability, and computing-efficiency. Basically, the details of BRBS can be found in [19].

3. Efficient rule-base modeling procedure to upgrade EBRBS as CBRBS

In this section, an efficient modeling procedure is proposed for CBRBS. First, the transformation of numeric data into belief distribution is discussed; second, new rule fusion and approximate rule weight calculation are proposed; finally, a comparison study is provided to prove the ability of CBRB modeling to overcome Challenge 1.

3.1. New function for numeric data transformation into belief distribution

The transformation of inputs into belief distributions is an important process in EBRBS and BRBS, and its role is the same to the fuzzification of FRBS. From the existing studies [19,23], numeric data is the most common input in system modeling and available transformation methods for numeric data mainly include matching function methods and utility-based transformation methods, in which the former one is required that the evaluation grades are characterized by fuzzy sets and the belief degree in the belief distribution is related to the fuzzy membership degree of a given numeric data; the latter one is required that the evaluation grades are characterized by utility values using utility-based transformation techniques [27]. Note that although the utility-based transformation method is an alternative way when fuzzy member functions are not available for numeric data transformation into belief distributions, they are the most commonly used transformation methods in the previous studies on EBRBS and BRBS.

According to [27], the utility-based transformation method can be defined as follows:

Definition 1 (*Adjacent Utility-based Transformation Function*)**.** Suppose $\{u(D_n);\ n = 1,\ldots, N\}$ is a set of utility values used for attribute D with $u(D_n) < u(D_{n+1})$ ($n = 1,\ldots, N$ - 1) and the kth numeric input of attribute D is y_k. Therefore, a belief distribution $S(D, y_k) = \{(D_n, \beta_n^k);\ n = 1,\ldots, N\}$ can be generated using the adjacent utilities of numeric input as follows:

$$\beta_n^k = \beta(y_k, D_n) = \begin{cases} \dfrac{y_k - u(D_{n-1})}{u(D_n) - u(D_{n-1})}, & \text{if } u(D_{n-1}) < y_k \le u(D_n) \text{ and } n > 1 \\ \dfrac{u(D_{n+1}) - y_k}{u(D_{n+1}) - u(D_n)}, & \text{if } u(D_n) \le y_k < u(D_{n+1}) \text{ and } n < N \\ 0, & \text{otherwise} \end{cases} \tag{8}$$

Based on Definition 1, the resulting belief distribution may be not suitable for some special situations. Taking pipeline leak detection problem [28] for example, suppose that attribute LS has three consequents $\{D_1 = Low, D_2 = Middle, D_3 = High\}$ and utility values $\{u(D_1), u(D_2), u(D_3)\} = \{0, 6, 8\}$. When the numeric data of LS is 4, the belief distribution therefore is $S(LS, 4) = \{(Low, 0.33), (Middle, 0.67), (High, 0)\}$, as shown in Fig. 3. However, the belief degree on *High* should be intuitively equal that on *Low* in oil pipeline leak detection because 4 is in the center between $u(High) = 8$ and $u(Low) = 0$.

Hence, a new function for the transformation of numeric data into belief distributions is proposed as follows:

Definition 2 (*Minimax Utility-based Transformation Function*)**.** Suppose $\{u(D_n);\ n = 1,\ldots, N\}$ is a set of utility values used for attribute D with $u(D_n) < u(D_{n+1})$ ($n = 1,\ldots, N$-1) and the kth numeric data of attribute D is y_k. Thus, a belief distribution $S(D, y_k) = \{(D_n, \beta_n^k);\ n = 1,\ldots, N\}$ can be generated using the minimum and maximum utilities as follows:

$$\beta_n^k = \beta(y_k, D_n) = \frac{\overline{\beta}(y_k, D_n)}{\sum_{t=1}^{N} \overline{\beta}(y_k, D_i)},$$
$$\overline{\beta}(y_k, D_n) = \begin{cases} \dfrac{y_k - u(D_1)}{u(D_n) - u(D_1)}, & \text{if } y_k < u(D_n) \\ 1, & \text{if } y_k = u(D_n) \\ \dfrac{u(D_N) - y_k}{u(D_N) - u(D_n)}, & \text{otherwise} \end{cases} \tag{9}$$

Based on Definition 2, the new transformation function provides a new perspective for the transformation of numeric data into belief distributions, in which the main advantage is that the obtained belief degrees can reflect the relationship between numeric data and utility values, so that it is possible to obtain belief distribution $\{(Low, 0.3), (Middle, 0.4), (High, 0.3)\}$ when the numeric data of LS is 4, as shown in Fig. 4.

3.2. Generation of cumulative belief rules based on information fusion

In EBRBS, an efficient rule generation method was applied to generate extended belief rules and its high computing-efficiency has been demonstrated in [23]. However, due to lack of rule reduction, the size of an EBRB would be infinitely increased with the increase of the number of available data. This is because the resulting EBRB satisfies the one-to-one correspondence between extended belief rules and input–output data pairs, namely Challenge 1 shown in Section 2.2. In order to overcome Challenge 1, a novel information fusion process is integrated into the EBRB modeling procedure for developing an efficient and effective CBRB modeling procedure. The corresponding detailed steps are as follows:

Step 1: Generation of an EBRB from input–output data pairs. Based on ***Step 1*** shown in Section 2.2, T extended belief rules R_k ($k = 1,\ldots, T$) can be generated from T input–output data pairs $<\boldsymbol{x_k}, y_k>$, in which each rule has a group of belief distribution $<S(U_i, x_{k,i}), S(D, y_k);\ i = 1,\ldots, M>$ ($S(U_i, x_{k,i}) = \{(A_{i,j}, \alpha_{i,j}^k);\ j = 1,\ldots, J_i\}$; $S(D, y_k) = \{(D_n, \beta_n^k);\ n = 1,\ldots, N\}$), and the rule weight of these T extended belief rules is assigned to be $\theta_k = 1$ according to Appendix A.

Step 2: Rule clustering of extended belief rules. Based on the maximum belief degree in belief distributions of each extended belief rule, multiple sets of similar extended belief rules can be obtained. For discussion purposes, a rule set related to M referential values A_{i,j_i}($j_i \in \{1,\ldots, J_i\}$; $i = 1,\ldots, M$) derived from M antecedent attributes is signed as $\boldsymbol{RC_{j_1\cdots j_M}}$. Hence, the corresponding set of extended belief rules can be expressed as follows:

$$R_k \in \boldsymbol{RC_{j_1\cdots j_M}};\ j_i = \arg \max_{j=1,\ldots,J_i} \{\alpha_{i,j}^k\};\ k = 1,\ldots,T;\ i = 1,\ldots,M \tag{10}$$

Step 3: Generation of CBRB based on the ER algorithm. For each rule set $\boldsymbol{RC_{j_1\cdots j_M}}$, all extended belief rules belonging to $\boldsymbol{RC_{j_1\cdots j_M}}$ should be used to generate a new extended belief rule $R_{j_1\cdots j_M}$, that called cumulate belief rule, and further forms a CBRB, as follows:

$$\begin{aligned} R_{j_1\cdots j_M}:\ & IF\ U_1\ is\ \{(A_{1,j}, \alpha_{1,j}^{j_1\cdots j_M});\ j = 1,\ldots,J_1\} \wedge \cdots \wedge U_M \\ & is\{(A_{M,j}, \alpha_{M,j}^{j_1\cdots j_M});\ j = 1,\ldots,J_M\}, \\ & THEN\ D\ is\ \{(D_n, \beta_n^{j_1\cdots j_M});\ n = 1,\ldots,N\},\ with\ \theta_{j_1\cdots j_M} \\ & and\ \{\delta_i;\ i = 1,\ldots,M\} \end{aligned} \tag{11}$$

where belief degree $\alpha_{i,j}^{j_1\cdots j_M}$ and $\beta_n^{j_1\cdots j_M}$ should be calculated by using the analytical ER algorithm to integrate the belief degree $\alpha_{i,j}^k$ and β_n^k of the extended belief rule R_k belongs to $\boldsymbol{RC_{j_1\cdots j_M}}$; rule weight $\theta_{j_1\cdots j_M}$ should be calculated by the rule weight of the extended belief rule R_k belongs to $\boldsymbol{RC_{j_1\cdots j_M}}$. More specifically, their calculation formulas are as follows:

$$\alpha_{i,j}^{j_1\cdots j_M} = ER_j(\alpha_{i,s}^k, \overline{\theta}_k, s = 1,\ldots,J_i, R_k \in \boldsymbol{RC_{j_1\cdots j_M}}) \tag{12}$$

$$\beta_n^{j_1\cdots j_M} = ER_n(\beta_s^k, \overline{\theta}_k, s = 1,\ldots,N, R_k \in \boldsymbol{RC_{j_1\cdots j_M}}) \tag{13}$$

$$\theta_{j_1\cdots j_M} = \frac{|\boldsymbol{RC_{j_1\cdots j_M}}|}{T} \tag{14}$$

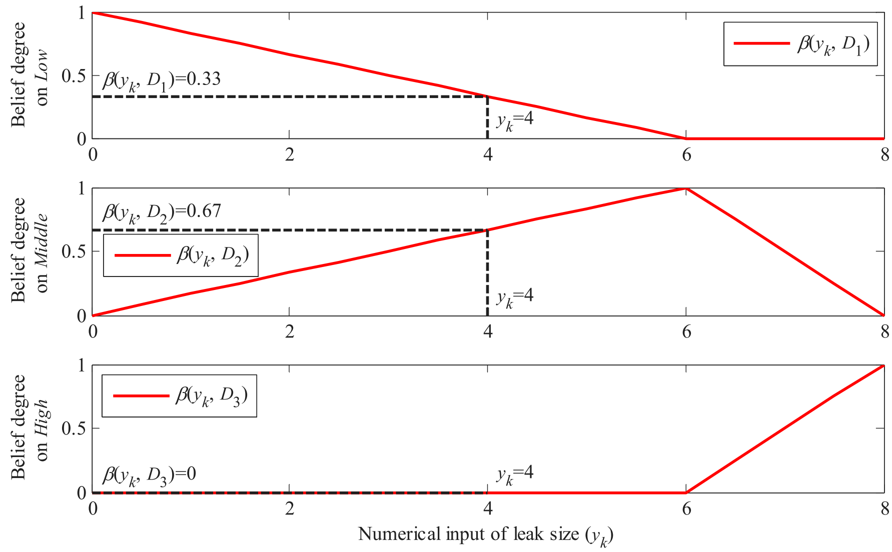


Fig. 3. Belief distribution obtained from adjacent utility-based transformation function.

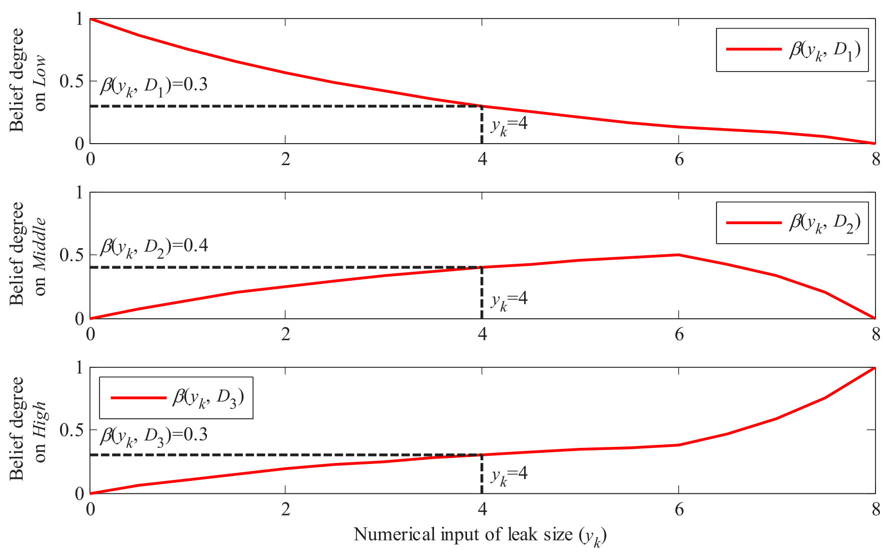


Fig. 4. Belief distribution obtained from minimax utility-based transformation function.

where $|\boldsymbol{RC}_{\boldsymbol{j}_1\cdots\boldsymbol{j}_M}|$ denotes the number of extended belief rules in rule set $\boldsymbol{RC}_{\boldsymbol{j}_1\cdots\boldsymbol{j}_M}$; $\overline{\theta}_k$ denotes the normalized rule weight of the kth extended belief rule in $\boldsymbol{RC}_{\boldsymbol{j}_1\cdots\boldsymbol{j}_M}$ and it can be calculated by

$$\overline{\theta}_k = \frac{1}{|\boldsymbol{RC}_{\boldsymbol{j}_1\cdots\boldsymbol{j}_M}|} \tag{15}$$

From above-mentioned three steps, the following remarks should be noted:

Remark 1. According to the new rule representation shown in Eq. (11), $R_{j_1\cdots j_M}$ can represent a special rule which has the maximum belief degrees in $A_{i,j_i}(i = 1,\ldots, M)$, namely $\alpha_{i,j_i}^{j_1\cdots j_M} > \alpha_{i,j}^{j_1\cdots j_M}(j = 1,\ldots, J_i;\ j\neq j_i;\ i = 1,\ldots, M)$, because it is obtained from the integration of the extended belief rules which all have the maximum belief degrees in A_{i,j_i} as well.

Remark 2. The upper limit of any CBRB's size is $\prod_{i=1}^{M} J_i$ on an minimax case that all extended belief rules have the maximum belief degrees on the combination of all referential values $\{A_{i,j};\ j = 1,\ldots, J_i\}$ for each antecedent attribute U_i ($i = 1,\ldots, M$). Additionally, a cumulative belief rule in Eq. (11) is the accumulation of multiple extended belief rules in Eq. (4) using Eqs. (12) to (15). *This makes sense that a CBRB is a cumulative version of an EBRB.*

3.3. First comparative analysis based on pipeline leak detection

In this subsection, the problem of oil pipeline leak detection [28] is re-examined to illustrate how BRBS, WM-FRBS, EBRBS, and CBRBS perform the procedure of rule-base modeling differently. In this problem, suppose that two antecedent attributes flow difference U_1 has 5, 6, 7, and 8 referential values $A_{1,j}$ ($j = 1,\ldots, J_1;\ J_1 = 5, 6, 7, 8$), respectively, and pressure difference U_2 has 4, 5, 6, and 7 referential values $A_{2,j}$ ($j = 1,\ldots, J_2;\ J_2 = 4, 5, 6, 7$), respectively. The utility values of these referential values are defined as follows:

$$u(A_{i,j}) = lb_i + (ub_i - lb_i)\frac{j-1}{J_i - 1}, i = 1, 2, j = 1, \ldots, J_i \tag{16}$$

where lb_i and ub_i denotes the lower and the upper bounds of the ith antecedent attribute. Based on the previous study [28], 500 data samples are collected in the three periods of 7 a.m. to 7:33

Table 1
Number of rules based on adjacent utility-based transformation function.

No. of rules in BRB/FRB/EBRB/CBRB		Flow difference			
		5 referential values	6 referential values	7 referential values	8 referential values
Pressure difference	4 referential values	20/8/500/8	24/14/500/14	28/14/500/14	28//15500/15
	5 referential values	25/10/500/10	30/17/500/17	35/16/500/16	40//16500/16
	6 referential values	30/12/500/12	36/21/500/21	42/20/500/20	48/20/500/20
	7 referential values	35/14/500/14	42/23/500/23	49/23/500/23	56/23/500/23

Table 2
Number of rules based on minimax utility-based transformation function.

No. of rules in BRB/FRB/EBRB/CBRB		Flow difference			
		5 referential values	6 referential values	7 referential values	8 referential values
Pressure difference	4 referential values	20/11/500/11	24/14/500/14	28/16/500/16	28/15/500/15
	5 referential values	25/14/500/14	30/16/500/16	35/20/500/20	40/18/500/18
	6 referential values	30/17/500/17	36/19/500/19	42/24/500/24	48/22/500/22
	7 referential values	35/19/500/19	42/21/500/21	49/27/500/27	56/26/500/26

a.m., 9:46 a.m. to 10:20 a.m. and 10:50 a.m. to 11:08 a.m. and all these 500 data are used to generate a BRB, FRB, EBRB, and CBRB, respectively.

As a result, the number of rules in the BRB, FRB, EBRB, and CBRB is shown in Tables 1 and 2 when using the adjacent utility-based and the minimax utility-based transformation functions, respectively.

From Tables 1 and 2, some conclusions can be summarized as follows:

(1) The number of rules in a BRB is related to the number of referential values for each antecedent attribute, and it is not related to the number of training data and the use of transformation functions, namely, $L = \prod_{i=1}^{M} J_i$;

(2) The number of rules in an EBRB is related to the number of training data, and it is not related to the number of referential values for each antecedent attribute and the use of transformation functions, namely, $L = T$;

(3) The number of rules in a FRB and a CBRB is related to the number of referential values used for each antecedent attribute and the use of transformation functions. Moreover, they have the same number of rules.

For the purpose of showing the rule distribution of BRB, FRB, EBRB, and CBRB, the rule is described by using the referential values in belief rules and fuzzy rules, as well as the referential values having the maximum belief degree in extended belief rules or cumulative belief rules, *i.e.*, the rules shown in Eqs. (2) and (3) can be described as < A_i^k; $i = 1,\ldots, M$> ($A_i^k \in \{A_{i,j};\ j = 1,\ldots, J_i\}$), and the rule shown in Eq. (4) can be described as < A_{i,j_i}; $j_i = \arg\max_{j=1,\ldots,J_i}\{\alpha_{i,j}^k\}$; $i = 1,\ldots, M$ >. Thus, when flow difference and pressure difference are assumed to have 8 and 7 referential values, respectively, namely negative large (NL), negative middle (NM), negative small (NS), zero (Z), positive small (PS), positive middle (PM), and positive large (PL), the rule distribution of BRB, FRBS, EBRB, and CBRB is shown in Figs. 5 and 6 for adjacent utility-based and minimax utility-based transformation functions.

From Figs. 5 and 6, some conclusions can be further summarized as follows:

(1) All rules in a BRB uniformly spread all over the combinations of all referential values for each antecedent attribute, and each combination only has one rule in adjacent utility-based and the minimax utility-based transformation functions.

(2) Comparing to the rule distribution of the BRB, the EBRB has different numbers of rules in the partial combinations of referential values for each antecedent attribute. This is because an EBRB is generated from training data directly.

(3) In the comparison of the rule distribution of a BRB, the similar point for the FRB and CBRB is that the latter two only have one rule in the combinations of all referential values for each antecedent attribute as well.

(4) By comparing with the rule distribution of the EBRB, the similar point for the FRB and CBRB is that their rule distributions are the same to the former one with different numbers of rules in each combination.

According to the above conclusions, there is a clear difference among BRB, EBRB, FRB, and CBRB, but too many similarities between FRB and CBRB. Hence, Fig. 7 provides a simple example when there are two belief distributions used for generating a fuzzy rule for WM-FRB and a cumulative belief rule for CBRB.

It is clear from Fig. 7 that a fuzzy rule is generated according to the linguistic labels which have the biggest important degree in the belief distribution, but a cumulative belief rule is generated by using all belief distributions. Hence, the former one always has to suffer from the information loss of the data which may be useful to decision making process; the latter one can keep all data in the rule with belief distributions.

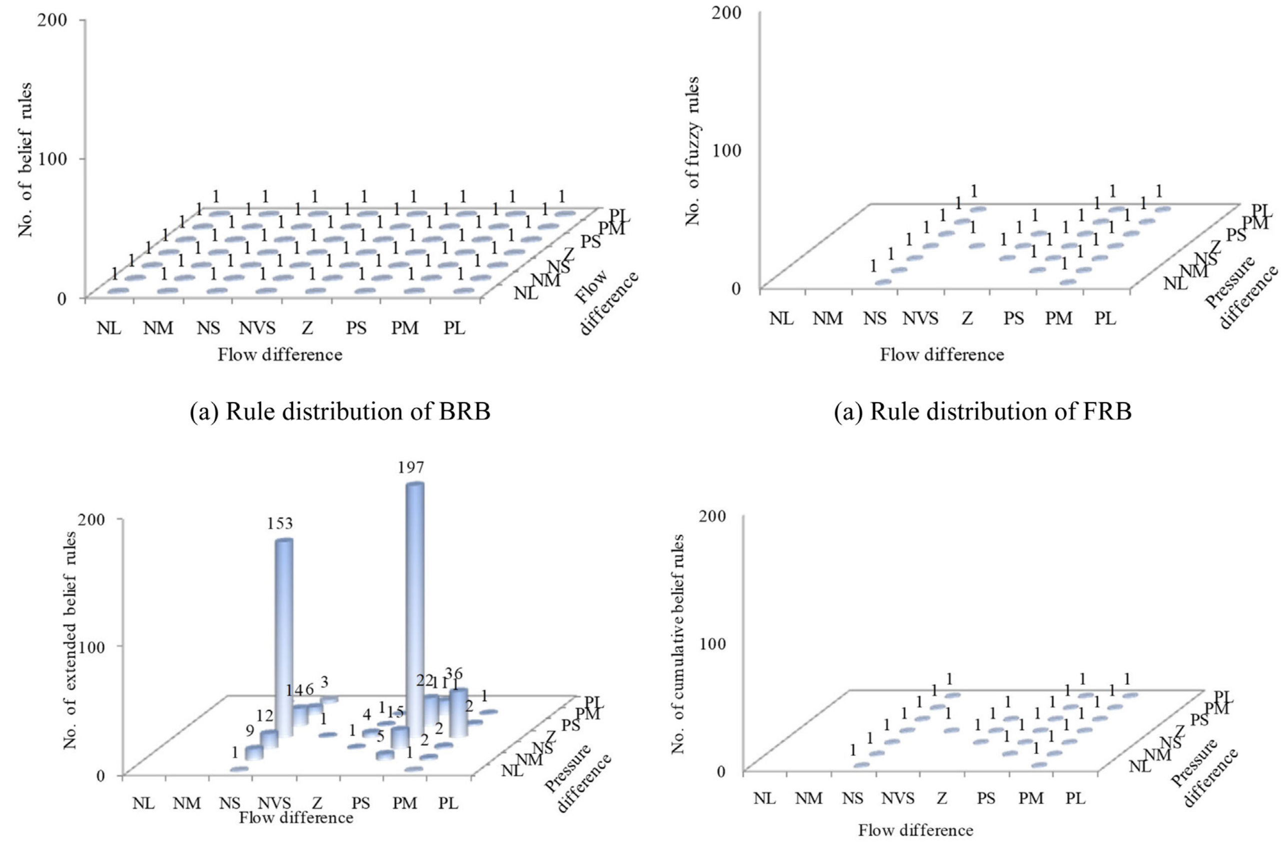
In summary, for the comparison of BRB, FRB, EBRB, CBRB modeling procedures, the experiment results have shown that the number of rules in a CBRB is far less than that of a BRB and an EBRB because the CBRB not only is based on training data to generate cumulative belief rule, instead of the combinations of all referential values for each antecedent attribute, but also involves rule fusion to integrate similar extended belief rules. Moreover, the CBRB has a powerful ability to ensure the information completeness of data better than the BRB and FRB. As a result, the CBRB modeling is able to overcome the rule boundlessness of EBRB modeling.

4. Efficient rule-base inference procedure to upgrade EBRBS as CBRBS

In this section, an efficient rule-base inference procedure is proposed for CBRBS. First, a nearest neighbor strategy is proposed to activate consistent cumulative rules; second, the integration of activated rules is developed to produce an output; finally, the comparison study is performed to prove the ability of CBRB inference procedure to overcome Challenge 2.

4.1. Nearest neighbor strategy to activate cumulative belief rules

The rule activation is one of the important processes in rule-based systems. Varying the formulation in this part of the systems can significantly affect the output produced. In an EBRBS, the rule activation is based on similarity measure between inputs and rule antecedents, so that all extended belief rules would be activated for replying any given input data, resulting in a negative influence



(a) Rule distribution of BRB (a) Rule distribution of FRB

(c) Rule distribution of EBRB (d) Rule distribution of CBRB

Fig. 5. Rule distribution for adjacent utility-based transformation function.

on EBRBS, namely Challenge 2 detailed in Section 2.2. In order to improve the rule inference process, a nearest neighbor rule activation strategy is investigated for the CBRBS as follows:

Suppose that an input data $\boldsymbol{x} = (x_1,\ldots, x_M)$ is provided for a CBRBS, which has a CBRB $\boldsymbol{R}$ with M sets of referential values $\{A_{i,j};\ j = 1,\ldots, J_i\}$ used for the ith $(i = 1,\ldots, M)$ antecedent attribute. In order to determine which cumulative belief rule needs to be activated, the activation priority of cumulative belief rule $R_{j_1\cdots j_M}(R_{j_1\cdots j_M} \in \boldsymbol{R})$, which has the maximum belief degree at $A_{i,j_i}(j_i\in\{1,\ldots, J_i\};\ i = 1,\ldots, M)$, should be calculated as follows:

$$p_{j_1\cdots j_M} = \max_{i=1,\ldots,M}\{p_i^{j_1\cdots j_M}\} \tag{17}$$

where $p_i^{j_1\cdots j_M}$ denotes the activation priority regarding the ith antecedent attribute and it is related to the distance between the referential value having the maximum belief degree in belief distribution $S(U_i, \boldsymbol{x}) = \{(A_{i,j}, \alpha_{i,j});\ j = 1,\ldots, J_i\}$ and A_{i,j_i}at the ith antecedent attribute. More specifically, $p_i^{j_1\cdots j_M}$ can be calculated by

$$p_i^{j_1\cdots j_M} = |\arg\max_{j=1,\ldots,J_i}\{\alpha_{i,j}\} - j_i| + 1 \tag{18}$$

Based on the obtained activation priority of all rules in $\boldsymbol{R}$, the cumulative belief rule, whose activation priority equals to the minimal activation priority derived from all cumulative belief rules, is selected as activated rules for replying input data $\boldsymbol{x}$. In other words, the set of activated rules can be expressed as follows:

$$\boldsymbol{AR}(\boldsymbol{x}) = \{R_{j_1\cdots j_M};\ p_{j_1\cdots j_M} = \min_{R_k\in\boldsymbol{R}}\{p_k\}\} \tag{19}$$

From above-mentioned strategy, the following remarks should be noted:

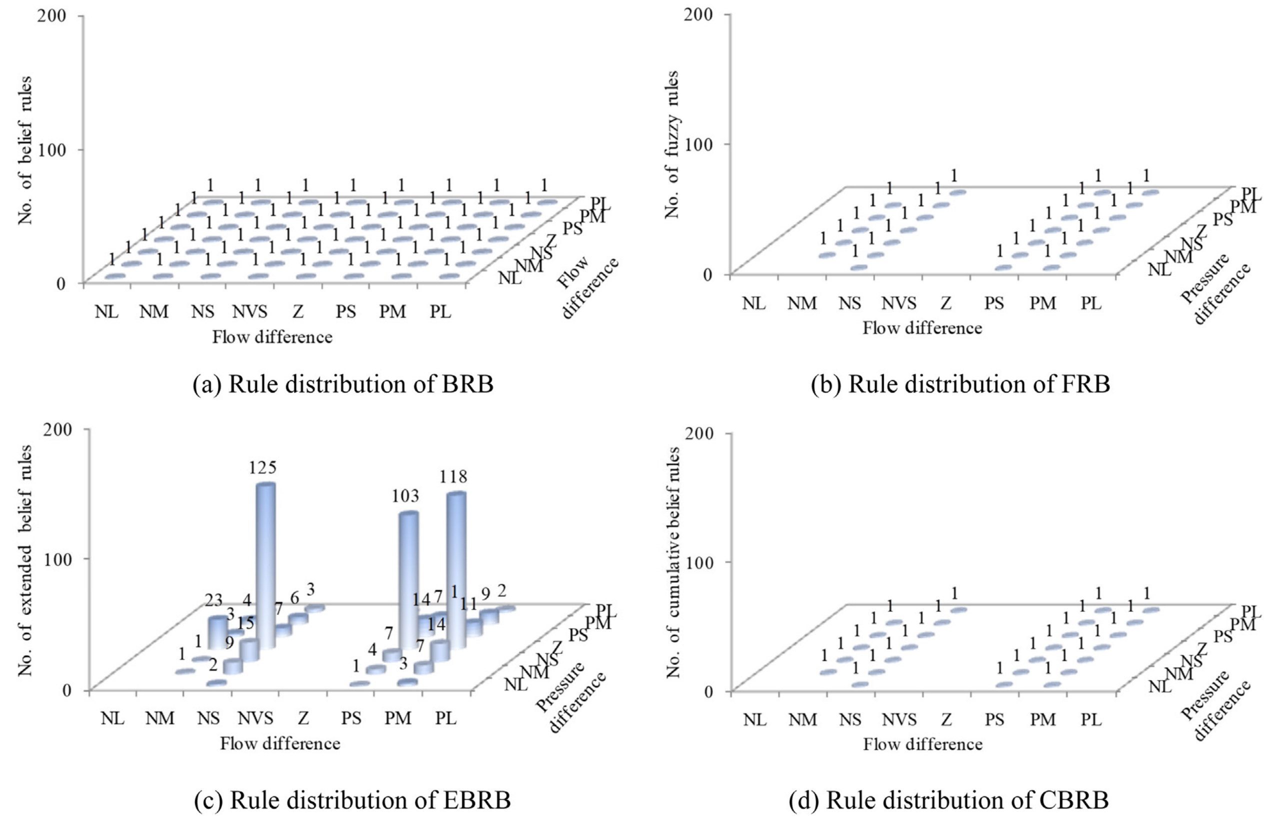
Remark 3. According to the determination of activated rules shown in Eq. (17), it is only to activate the cumulative belief rules which have the minimum activation priority that also means a maximum similarity between inputs and rules. Hence, the proposed rule activation strategy can activate consistent cumulative belief rules to reply any given input data.

Remark 4. Any activated rule in CBRB can represent an integrated rule obtained from the pre-integration of the extended belief rules which have same rule antecedent in EBRB using the analytical ER algorithm (that is also the inference engine of a CBRBS). Hence, *this makes sense that a CBRBS is a cumulative version of an EBRBS*.

In order to graphically illustrate how to activate cumulative belief rules, assume there are 2 antecedent attributes U_i $(i = 1, 2)$ with 5 referential values $A_{i,j}$ $(j = 1,\ldots, 5)$ for each antecedent attribute, and 15 cumulative belief rules $\{R_{1,1}, R_{1,3}, R_{1,4}, R_{2,2}, R_{2,3}, R_{2,5}, R_{3,1}, R_{3,4}, R_{3,5}, R_{4,1}, R_{4,4}, R_{5,1}, R_{5,2}, R_{5,3}, R_{5,5}\}$ in a CBRB, where $R_{j_1,j_2}(j_i \in \{1,\ldots,J_i\};\ i = 1,2)$ denotes a cumulative belief rule having the maximum belief degree in referential values A_{i,j_i}. For the purpose of illustrating the rule distribution of CBRB, cumulative belief rule R_{j_1,j_2} are described by < A_{i,j_1}, A_{i,j_2}> and they therefore can be listed in Fig. 8. Similarly, assume that there are two input data $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$, which are also described by the referential values that have the maximum belief degree in Fig. 8 after they are transformed into belief distributions.

From Fig. 8, two situations of activating cumulative belief rules can be discussed as follows:

(1) When input data $\boldsymbol{x}_1$ locates in the place shown in Fig. 8(a), the activation priority of each cumulative belief rule is $p_{1,1} =$



(a) Rule distribution of BRB (b) Rule distribution of FRB

(c) Rule distribution of EBRB (d) Rule distribution of CBRB

Fig. 6. Rule distribution for minimax utility-based transformation function.

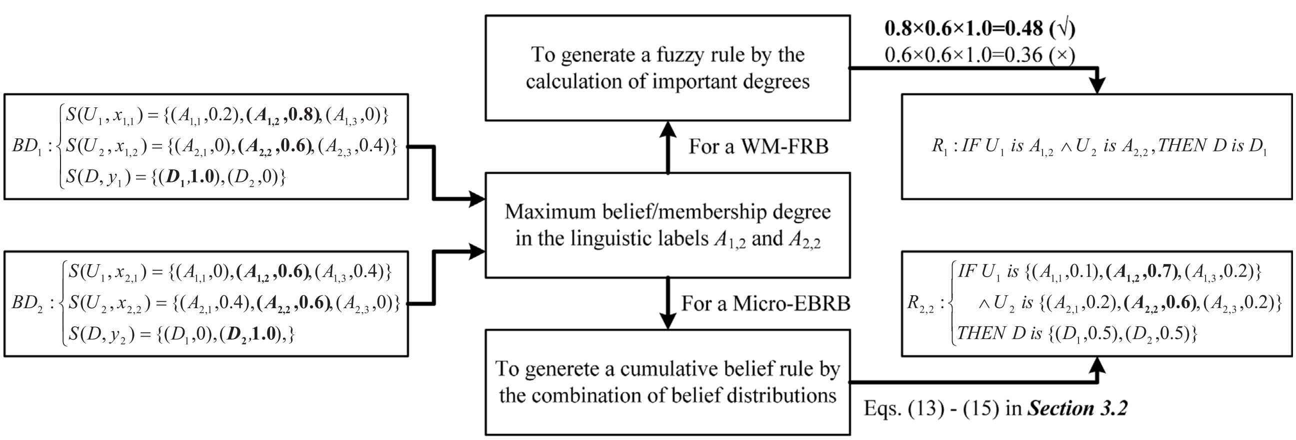


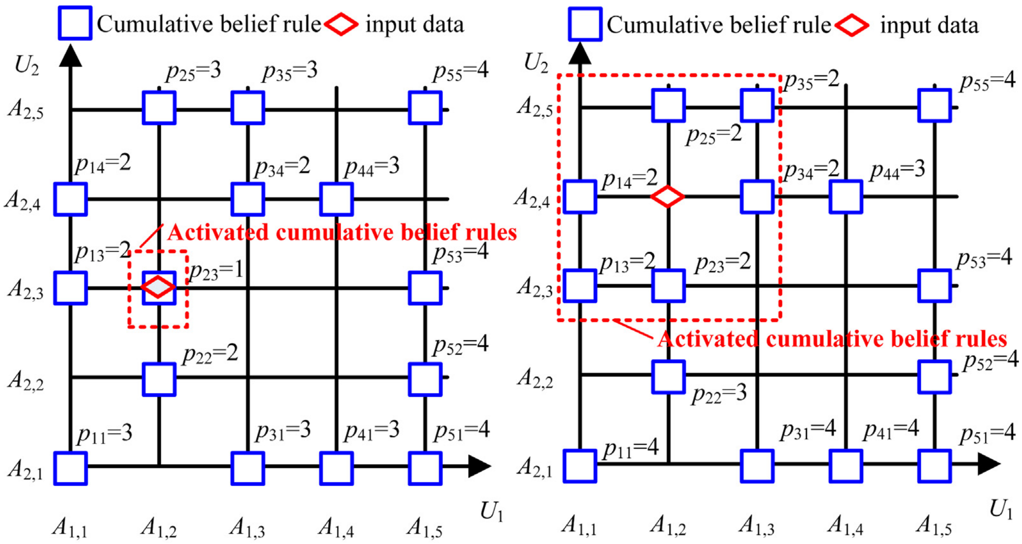
Fig. 7. Rule generation in WM-FRB and CBRB.

max{3, 2} = 3, $p_{1\cdot 3}$ = max{1, 2} = 2, $p_{1\cdot 4}$ = max{2, 2} = 2, $p_{2\cdot 2}$ = max{2, 1} = 2, $p_{2\cdot 3}$ = max{1, 1} = 1, $p_{2\cdot 5}$ = max{3, 1} = 3, $p_{3\cdot 1}$ = max{3, 2} = 3, $p_{3\cdot 4}$ = max{2, 2} = 2, $p_{3\cdot 5}$ = max{3, 2} = 3, $p_{4\cdot 1}$ = max{3, 3} = 3, $p_{4\cdot 4}$ = max{2, 3} = 3, $p_{5\cdot 1}$ = max{3, 4} = 4, $p_{5\cdot 2}$ = max{2, 4} = 4, $p_{5\cdot 3}$ = max{1, 4} = 4, and $p_{5\cdot 5}$ = max{3, 4} = 4. Hence, the set of activated rules $\boldsymbol{AR}(\boldsymbol{x_1})$ only includes $R_{2,3}$ because of $p_{2\cdot 3} = 1 = \min\{ p_{1\cdot 1}, p_{1\cdot 3}, p_{1\cdot 4}, p_{2\cdot 2}, p_{2\cdot 3}, p_{2\cdot 5}, p_{3\cdot 1}, p_{3\cdot 4}, p_{3\cdot 5}, p_{4\cdot 1}, p_{4\cdot 4}, p_{5\cdot 1}, p_{5\cdot 2}, p_{5\cdot 3}, p_{5\cdot 5}\}$.

(2) When input data $\boldsymbol{x_2}$ locates in the place shown in Fig. 8(b), the activation priority of each cumulative belief rule is $p_{1\cdot 1}$ = max{4, 2} = 4, $p_{1\cdot 3}$ = max{2, 2} = 2, $p_{1\cdot 4}$ = max{1, 2} = 2, $p_{2\cdot 2}$ = max{3, 1} = 3, $p_{2\cdot 3}$ = max{2, 1} = 2, $p_{2\cdot 5}$ = max{2, 1} = 2, $p_{3\cdot 1}$ = max{4, 2} = 4, $p_{3\cdot 4}$ = max{1, 2} = 2, $p_{3\cdot 5}$ = max{2, 2} = 2, $p_{4\cdot 1}$ = max{4, 3} = 4, $p_{4\cdot 4}$ = max{1, 3} = 3, $p_{5\cdot 1}$ = max{4, 4} = 4, $p_{5\cdot 2}$ = max{3, 4} = 4, $p_{5\cdot 3}$ = max{2, 4} = 4, and $p_{5\cdot 5}$ = max{2, 4} = 4. Hence, the set of activated rules $\boldsymbol{AR}(\boldsymbol{x_2})$ includes $R_{1,3}$, $R_{1,4}$, $R_{2,3}$, $R_{2,5}$, $R_{3,4}$, and $R_{3,5}$ because of $p_{1\cdot 3} = p_{1\cdot 4} = p_{2\cdot 3} = p_{2\cdot 5} = p_{3\cdot 4} = p_{3\cdot 5} = 2 = \min\{ p_{1\cdot 1}, p_{1\cdot 3}, p_{1\cdot 4}, p_{2\cdot 2}, p_{2\cdot 3}, p_{2\cdot 5}, p_{3\cdot 1}, p_{3\cdot 4}, p_{3\cdot 5}, p_{4\cdot 1}, p_{4\cdot 4}, p_{5\cdot 1}, p_{5\cdot 2}, p_{5\cdot 3}, p_{5\cdot 5}\}$.

4.2. Integration of activated rules to produce inference outputs

After activating cumulative belief rules for the given input data, all these activated rules should be integrated to produce inference outputs. In this process, the calculation of similarity between belief distributions is fundamental, since it is the basis of the calculation of matching degrees between inputs and rule antecedents and further affects the performance of the entire system. From the study on [23], the method used to measure the similarity between belief distributions is based on Euclidean distance as follows:



(a) First situation (b) Second situation

Fig. 8. Two situations of activating cumulative belief rules.

Definition 3 (*Euclidean Distance-based Similarity Measure*)**.** Suppose there are two belief distributions $S(U_i, x_{k,i}) = \{(A_{i,j}, \alpha_{i,j}^k); j = 1,\ldots, J_i\}$ and $S(U_i, x_i) = \{(A_{i,j}, \alpha_{i,j}); j = 1,\ldots, J_i\}$ in antecedent attribute U_i, thus the similarity of belief distributions should be calculated by:

$$Sim(U_i, x_i, x_{k,i}) = 1 - \sqrt{\sum_{j=1}^{J_i}(\alpha_{i,j}^k - \alpha_{i,j})^2} \tag{20}$$

However, because of the deficiencies in Definition 3, *e.g.*, counterintuitive similarity $Sim(U_i, x_i, x_{k,i})<0$ and insensitive to referential values $A_{i,j}$, the following revised similarity measure methods [35] were developed from Definition 3:

Definition 4 (*Standardized Euclidean Distance (SED)-based Similarity Measure*)**.** On the same situation of Definition 3, the standardized Euclidean distance-based similarity for $S(U_i, x_{k,i})$ and $S(U_i, x_i)$ can be calculated by:

$$Sim(U_i, x_i, x_{k,i}) = 1 - \sqrt{\frac{\sum_{j=1}^{J_i}(\alpha_{i,j}^k - \alpha_{i,j})^2}{2}} \tag{21}$$

Definition 5 (*Constrained Euclidean Distance (CED)-based Similarity Measure*)**.** On the same situation of Definition 3, the constrained Euclidean distance-based similarity for $S(U_i, x_{k,i})$ and $S(U_i, x_i)$ can be calculated by:

$$Sim(U_i, x_i, x_{k,i}) = 1 - \min\{1, \sqrt{\sum_{j=1}^{J_i}(\alpha_{i,j}^k - \alpha_{i,j})^2}\} \tag{22}$$

Definition 6 (*Weighted Euclidean Distance (WED)-based Similarity Measure*)**.** On the same situation of Definition 3, when utility values $\{u(A_{i,j}); j = 1,\ldots, J_i\}$ are provided for referential values $\{A_{i,j}; j = 1,\ldots, J_i\}$, thus the weighted Euclidean distance-based similarity for $S(U_i, x_{k,i})$ and $S(U_i, x_i)$ can be calculated by:

$$Sim(U_i, x_i, x_{k,i}) = 1 - \sqrt{\frac{\sum_{j=1}^{J_i}\varepsilon_{i,j}(\alpha_{i,j}^k - \alpha_{i,j})^2}{\varepsilon_{N-1} + \varepsilon_N}} \tag{23}$$

where $\varepsilon_{i,j}$ denotes the weight of referential value $A_{i,j}$ and it is calculated by

$$\varepsilon_{i,j} = e^s,\ s = \frac{u(A_{i,j}) - u(A_{i,1})}{u(A_{i,J_i}) - u(A_{i,1})} \tag{24}$$

On the bases of the above three improved similarity measure functions, the integration of activated rules to produce inference outputs for replying given input data is introduced as follows:

Step 1: Calculation of activation weight for each activated rule. Suppose that a set of activated rules $\boldsymbol{AR}(\boldsymbol{x})$ related to input data $\boldsymbol{x}$ $= (x_i; i = 1,\ldots, M)$ is obtained from the nearest neighbor strategy shown in Section 4.1. Hence, for any activated rule $R_{j_1\cdots j_M}(R_{j_1\cdots j_M} \in \boldsymbol{AR}(\boldsymbol{x}))$, its activation weight, denoted as $w_{j_1\cdots j_M}$, can be calculated as follows:

$$w_{j_1\cdots j_M} = \frac{\theta_{j_1\cdots j_M}\prod_{i=1}^{M} Sim(U_i, x_i, R_{j_1\cdots j_M})^{\bar{\delta}_i}}{\sum_{R_k}^{AR(\mathbf{x})}\theta_k\prod_{i=1}^{M} Sim(U_i, x_i, R_k)^{\bar{\delta}_i}},\ \bar{\delta}_i = \frac{\delta_i}{\max_{i=1,\ldots,M}\{\delta_i\}} \tag{25}$$

where $\theta_{j_1\cdots j_M}$ is the weight of rule $R_{j_1\cdots j_M}$; δ_i is the weight of antecedent attribute U_i; $Sim(U_i, x_i, R_{j_1\cdots j_M})$ is the similarity between $S(U_i, x_i)$ derived from x_i and $S(U_i, R_{j_1\cdots j_M})$ derived from $R_{j_1\cdots j_M}$.

Step 2: Integration of all activated rules to reply input data $\boldsymbol{x}$. After calculating activation weights, all activated rules should be integrated using the analytical ER algorithm as follows:

$$\beta_n = ER_n(\beta_s^{j_1\cdots j_M}, w_{j_1\cdots j_M}, s = 1,\ldots,N, R_{j_1\cdots j_M} \in AR(\boldsymbol{x})) \tag{26}$$

Afterwards, for regression problems, assume that $u(D_n)$ represents utility value of consequent D_n at attribute D and satisfies $u(D_1)\leq u(D_2)\leq\ldots \leq u(D_N)$, the inference output is then described as:

$$f(\boldsymbol{x}) = \sum_{n=1}^{N} u(D_n)\beta_n + (1 - \sum_{n=1}^{N}\beta_n)\frac{u(D_1) + u(D_N)}{2} \tag{27}$$

For classification problems, suppose that D_n represents the nth class, the inference output is then described as:

$$f(\boldsymbol{x}) = D_t,\ t = \arg\max_{n=1,\ldots,N}\{\beta_n\} \tag{28}$$

Table 3
Average activation times per rule under adjacent utility-based transformation function.

Average activation times per rule in BRB/ FRB/ EBRB/ CBRB		Flow difference			
		5 referential values	6 referential values	7 referential values	8 referential values
Pressure difference	4 referential values	99.2/125.0/500/62.5	83.3/137.2/500/35.7	71.4/137.6/500/35.7	62.5/112.1/500/33.3
	5 referential values	51.8/65.2/500/50.0	43.5/73.4/500/29.4	37.3/77.3/500/31.3	32.6/68.2/500/31.3
	6 referential values	65.5/82.5/500/41.7	55.0/91.6/500/23.8	47.1/96.1/500/25.0	41.3/82.5/500/25.0
	7 referential values	37.0/46.6/500/35.7	31.0/53.3/500/21.7	26.6/54.3/500/21.7	23.3/48.0/500/21.7

Table 4
Average activation times per rule under minimax utility-based transformation function.

Average activation times per rule in BRB/ FRB/ EBRB/ CBRB		Flow difference			
		5 referential values	6 referential values	7 referential values	8 referential values
Pressure difference	4 referential values	500/500/500/45.5	500/500/500/35.7	500/500/500/31.3	500/500/500/33.3
	5 referential values	500/500/500/35.7	500/500/500/31.3	500/500/500/25.0	500/500/500/27.8
	6 referential values	500/500/500/29.4	500/500/500/26.3	500/500/500/20.8	500/500/500/22.7
	7 referential values	500/500/500/26.3	500/500/500/23.8	500/500/500/18.5	500/500/500/19.2

4.3. Second comparative analysis based on pipeline leak detection

In this subsection, the problem of pipeline leak detection is re-examined to illustrate how BRBS, WM-FRBS, EBRBS, and CBRBS perform the rule-base inference procedure differently. In this problem, suppose that the antecedent attribute flow difference has 5, 6, 7, and 8 referential values $A_{1,j}$ ($j = 1,\ldots, J_1$; $J_1 = 5, 6, 7, 8$), respectively, and the antecedent attribute pressure difference has 4, 5, 6, and 7 referential values $A_{2,j}$ ($j = 1,\ldots, J_2$; $J_2 = 4, 5, 6, 7$), respectively. The utility values of the two sets of referential values are shown in Eq. (16). Furthermore, 500 data samples are collected in the three periods of 7 a.m. to 7:33 a.m., 9:46 a.m. to 10:20 a.m. and 10:50 a.m. to 11:08 a.m. and all 500 data are used to generate BRB, FRB, EBRB, and CBRB, and also used to validate the rule inference of BRBS, WM-FRBS, EBRBS, and CBRBS.

As a result, the average activation times per rule in the BRB, FRB, EBRB, and CBRB inference are shown in Tables 3 and 4 when using the adjacent utility-based and the minimax utility-based transformation functions.

From Tables 3 and 4, some conclusions can be summarized as follows:

(1) The average activation times per rule in the EBRB inference are 500, which indicate that all extended belief rules in the EBRB have to be activated to reply any given input data, leading to the rule inconsistency of the EBRB inference finally. In addition, it is clear from Tables 3 and 4 that the transformation function has no influence on the average activation times per rule in the EBRB inference.

(2) For the adjacent utility-based transformation function, an overall trend of BRB, FRB, and CBRB inference is that average activation times per rule decrease with the increase of number of referential values. This is because more sophisticated rule distribution can be generated when providing more referential values and hereby more accurate but fewer rules should be activated. Meanwhile, a clear distinction comparing to the EBRB inference is that the number of activation rules is a part of all rules in these three rule bases when replying a given input data.

(3) For the minimax utility-based transformation function, the average activation times per rule for the BRB and FRB inference equal those for the EBRB inference, which indicates that the BRB and FRB inferences have to suffer from the problem of rule inconsistency. This is because the belief distribution generated by the minimax utility-based transformation function does not have zero belief degree at any referential value and hereby all rules in the BRB and FRB should be activated for replying any given input data.

For the purpose of further comparing with BRB, FRB, EBRB, and CBRB inference procedures, the rule distributions shown in Figs. 5 and 6 are introduced when antecedent attribute flow difference has 8 referential values, namely NL, NM, NS, NVS, Z, PS, PM, and PL, and antecedent attribute pressure difference has 7 referential values, namely NL, NM, NS, Z, PS, PM, and PL. The corresponding rule activation times per rule of these four rule-bases can be found in Figs. 9 and 10.

From Figs. 9 and 10, some conclusions can be further summarized as follows:

(1) Although all rules in a BRB uniformly spread all over the combinations of all referential values for each antecedent attribute, the adjacent utility-based transformation function would cause the zero rule activation problem that some of belief rules does not be activated for replying any input data. Accordingly, the minimax utility-based transformation function would cause the rule inconsistency problem that all belief rules are activated for the same input data.

(2) Although FRB, EBRB, CBRB have the same rule distribution to each other, only the CBRB inference not only can avoid the zero rule activation problem, but also is able to overcome the rule inconsistency problem in the adjacent utility-based and minimax utility-based transformation functions. This is because the proposed rule activation strategy in Section 4.1 can effectively activate cumulative belief rules.

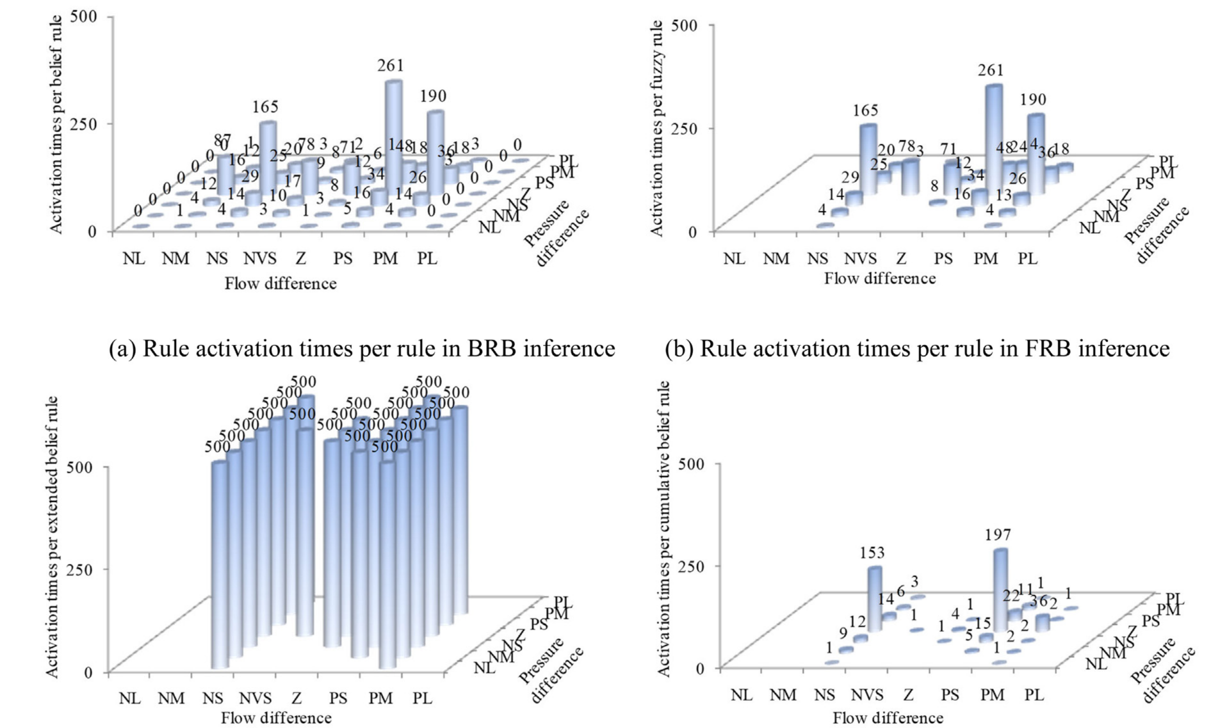
In summary, for the comparison of BRB, FRB, EBRB, CBRB inference procedures, the experiment results have shown that CBRBS can activate reasonable rules to reply any given input data better than BRBS, WM-FRBS, and EBRBS, so that CBRB inference is able to overcome the rule inconsistency problem of EBRB inference outlined in Section 2.2.

5. Generic CBRBS and it properties

In this section, the basic framework of generic CBRBS is provided firstly to facilitate its appropriate use, followed by the analysis of inherent properties and the proof of universal approximation property.

5.1. Basic framework of CBRBS

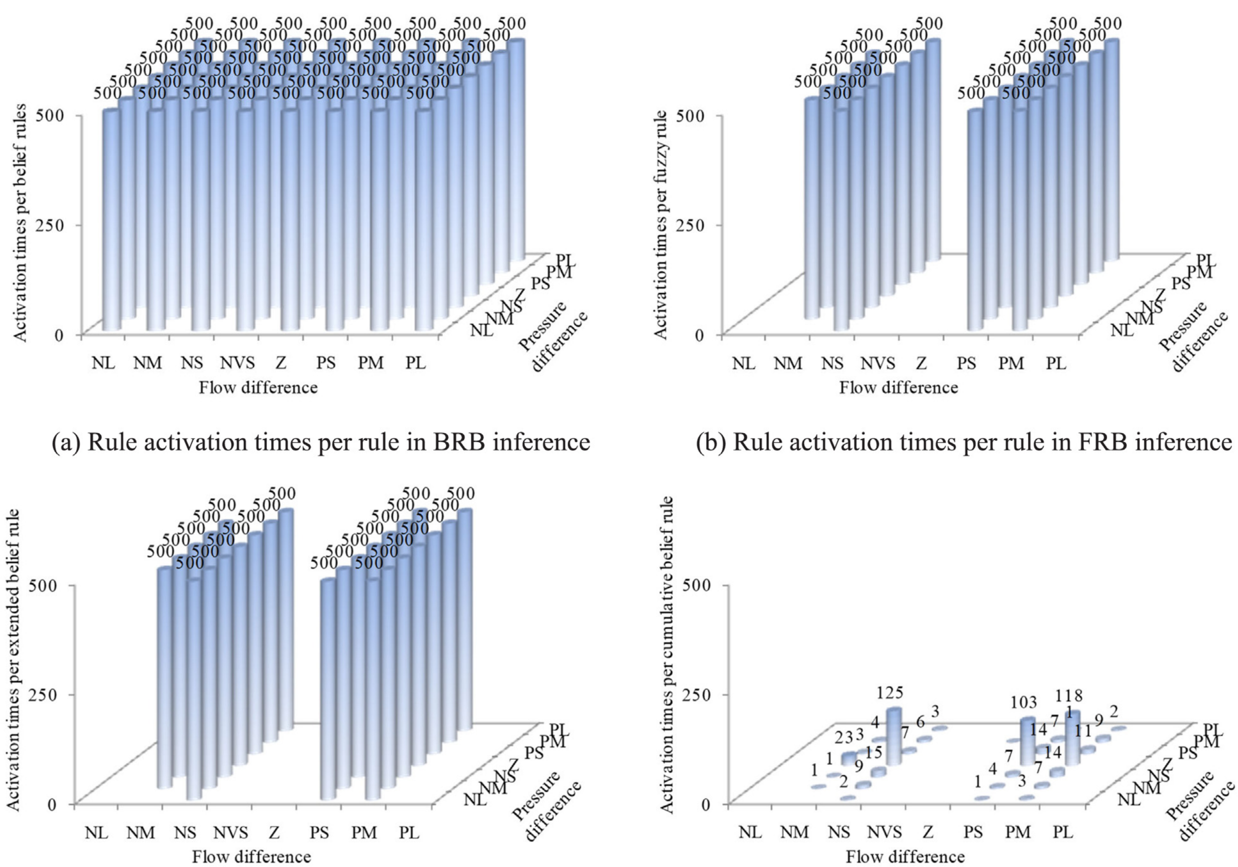
In order to clearly illustrate how a CBRBS work efficiently and effectively, based on the efficient rule-base modeling and inference procedures proposed in Sections 3 and 4, the basic framework of CBRBS is proposed in this section and its details are shown in Fig. 11. From Fig. 11, the CBRBS consists of two components: CBRB modeling procedure and CBRB inference procedure, in which the former is the process of generating a CBRB from basic parameters and training data; the latter is the process of rule-base inference using the CBRB.



(a) Rule activation times per rule in BRB inference (b) Rule activation times per rule in FRB inference

(c) Rule activation times per rule in EBRB inference (d) Rule activation times per rule in CBRB inference

Fig. 9. Rule activation times per rule under adjacent utility-based transformation function.



(a) Rule activation times per rule in BRB inference (b) Rule activation times per rule in FRB inference

(c) Rule activation times per rule in EBRB inference (d) Rule activation times per rule in CBRB inference

Fig. 10. Rule activation times per rule under minimax utility-based transformation function.

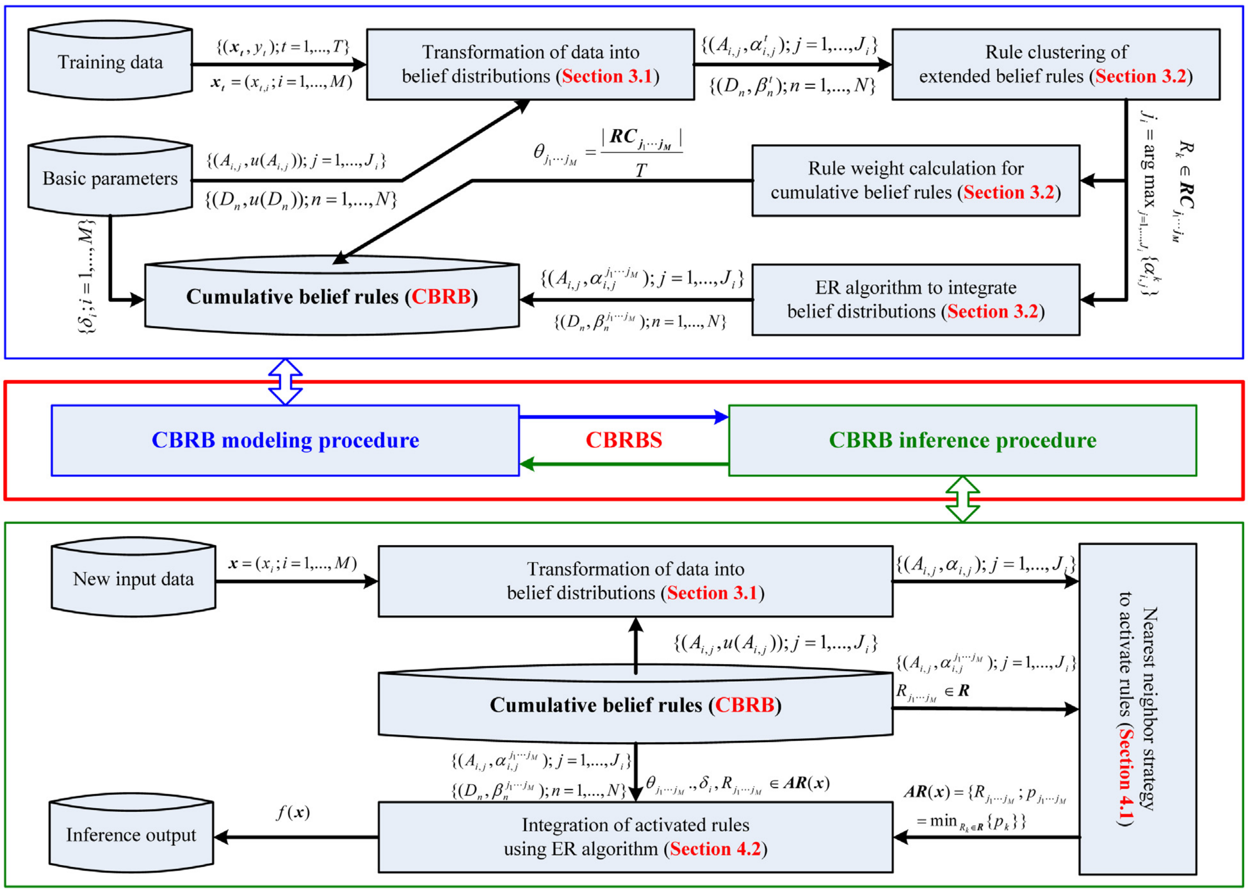


Fig. 11. Basic framework of CBRBS.

For the CBRB modeling procedure, it consists of the following steps and the corresponding pseudocode is provided in Appendix B:

Step 1: Transformation of inputs into belief distributions. According to the given training data and basic parameters, the functions discussed in Section 3.1 can be alternative ways to transform numeric data into belief distributions. Actually, the training data are not necessary to be numeric data; they could be a belief distribution of qualitative referential values.

Step 2: Rule clustering of extended belief rules. The purpose of this step is to reduce the number of extended belief rules by clustering the rules, which have the maximum belief degree in the same referential value, into a rule group. The formulaic rule clustering is in Eq. (10) at Section 3.2.

Step 3: Information fusion to generate cumulative belief rules. In this step, a cumulative belief rule should be generated in the form of each rule group obtained from ***Step 2***, in which the belief distributions and rule weight of each cumulative belief rule can be calculated by using Eqs. (12) to (14) in Section 3.2.

For the CBRB inference procedure, it consists of the following steps and the corresponding pseudocode is provided in Appendix C:

Step 1: Transformation of new inputs into belief distributions. For any given new input, the first step is to transform the input into belief distributions which is the same to the first step of the CBRB modeling procedure. Also, the new inputs are not necessary to be numeric data; they could be a belief distribution of qualitative referential values.

Step 2: Nearest neighbor strategy to activate rules. The purpose of this step is to activate rules based on the belief distribution of the new input data and cumulative belief rules, so that consistent cumulative belief rules can be activated for replying the new input data. The formulaic rule activation is in Eqs. (17) to (19) in Section 4.1.

Step 3: Integration of activated rules using the ER algorithm. In this step, an activation weight is calculated for each activated rule and thereafter all activated rules are integrated using Eq. (26) in Section 4.2. Finally, an inference output can be obtained according to classification or regression problems.

Remark 5. In the CBRB modeling procedure, the basic parameters, including utility values and attribute weights, can be provided based on expert knowledge. However, the subjective of expert knowledge may have negative influence on CBRBS. Hence, the following parameter learning model can be used to optimize the basic parameters of CBRB:

$$Min \sum_{t=1}^{T} G(y_t, f(\boldsymbol{x}_t)) \tag{29a}$$

$$s.t. 0 < \delta_i \leq 1; i = 1, \ldots, M \tag{29b}$$

$$u(A_{i,j}) \leq u(A_{i,j+1}); i = 1, \ldots, M; j = 1, \ldots, J_i - 1 \tag{29c}$$

$$u(A_{i,1}) = lb_i, u(A_{i,J_i}) = ub_i; i = 1, \ldots, M \tag{29d}$$

Table 5
Computing complexity analysis of the CBRB modeling.

Modeling Procedure	Computing complexity	Explanations
Step 1	$O(T \times \sum_{i=1}^{M} J_i + T \times N)$	Transformation of T inputs into $T \times M$ belief distributions for the M antecedent attributes and T belief distributions for the one consequent attribute.
Step 2	$O(T \times \sum_{i=1}^{M} J_i)$	Searching for the maximum belief degree from $T \times M$ belief distributions to cluster the extended belief rules.
Step 3	$O(T \times \sum_{i=1}^{M} J_i + T \times N)$	Integration of $T \times M$ belief distributions in the M antecedent attributes and T belief distributions in the one consequent attribute.
Step 4	$O(T)$	Calculation of rule weights for T cumulative belief rules at most by counting the total number of extended belief rules in the same rule group.

Table 6
Computing complexity analysis of the CBRB Inference.

Inference Procedure	Computing complexity	Explanations
Step 1	$O(\sum_{i=1}^{M} J_i)$	Transformation of one new input into M belief distributions for the M antecedent attributes.
Step 2	$O(L \times \sum_{i=1}^{M} J_i)$	Searching for the maximum belief degree from $L \times M$ belief distributions to determine activation priorities for the L rules.
Step 3	$O(L \times \sum_{i=1}^{M} J_i + L \times N)$	Measuring the similarity of $L \times M$ belief distributions in the M antecedent attributes and integrating L belief distributions in the consequent attribute.

$$u(D_n) \leq u(D_{n+1});\ n = 1, \ldots, N-1 \tag{29e}$$

$$u(D_1) = lb, u(D_N) = ub \tag{29f}$$

where lb_i and ub_i are the lower and upper bounds of the ith antecedent attribute; lb and ub are the lower and upper bounds of consequent attribute. $f(\boldsymbol{x_t})$ is the inference output of CBRBS for input data $\boldsymbol{x_t}$, y_t is the real output of $\boldsymbol{x_t}$. In addition, the target function $G(y_t, f(\boldsymbol{x_t}))$ can be error for regression problems or accuracy for classification problems.

5.2. Inherent properties of the CBRBS

In this section, the inherent properties of the CBRBS, including linear computing complexity, boundedness, and continuity properties, are studies. It is worth noting that when the utility value of each consequent is provided for a CBRBS, an inference output can be produced in both regression and classification problems. Hence, the inference output in Eq. (27) is regarded as the inference output of CBRBS when analyzing its inherent properties.

Lemma 1 (*Linear Computing Complexity Property*)**.** *For the given number of referential values and consequents in a CBRBS, the computing complexity of the CBRB modeling and inference procedures is linear.*

Proof. In a CBRBS, assume that there are J_i ($i = 1,\ldots, M$) referential values for M antecedent attributes, N consequents for one consequent attribute, L extended belief rules generated from T training data for the resulting CBRB. Based on the steps shown in Section 5.1, Tables 5 and 6 analyze and provide the computing complexity of CBRB modeling and inference procedures. It is clear from Tables 5 and 6 that the final computing complexity of CBRB modeling and inference procedures is $O(T \times \sum_{i=1}^{M} J_i + T \times N)$ and $O(L \times \sum_{i=1}^{M} J_i + L \times N)$, respectively. Hence, for given number of referential values and consequents, it is proved that a CBRBS has a linear computing complexity for CBRB modeling and inference procedures.

From Lemma 1, the high-efficiency of the CBRBS is demonstrated and measured by using the computing complexity analysis shown in Tables 5 and 6. Furthermore, a corollary is further provided to illustrate the characteristic of CBRBS.

Corollary 1. *A CBRBS is able to scale up to big data problems*

Proof. It has been proved in Lemma 1 that a CBRBS has linear computing complexity in the process of rule-base modeling and inference, which is important prerequisite and guarantee to ensure that a system/approach can be used to handle big data problems. As a result, it only needs to prove the parallelization capability of the CBRBS.

On the one hand, according to the CBRB modeling procedure, it is obvious that each input–output data pair can generate a set of belief distributions and then determine its corresponding rule set under considered as independent each other. Afterwards, each rule set can also independently generate a cumulative belief rule and its rule weight. On the other hand, based on the CBRB inference procedure, there is no association between the inference procedures of a CBRBS to reply different new input data. Hence, a CBRBS can be easy implemented by a parallelization way. Consequently, it is proved that a CBRBS is able to scale up to big data problems.

According to Corollary 1, a discussion of CBRBS, EBRBS, BRBS, and WM-FRBS is provided to show which one has a better efficiency to deal with big data problems. Firstly, based on the previous study [24], the computing complexity of EBRBS modeling and inference procedures is $O(T^2 \times \sum_{i=1}^{M} J_i + T^2 \times N)$ and $O(T \times \sum_{i=1}^{M} J_i + T \times N)$, which means that EBRBS has a worse efficiency in both modeling and inference procedures compared to CBRBS. Secondly, because of the fact that BRB modeling procedure is highly dependent on time-consuming parameter learning process [20], BRBS usually is regarded as the least desirable approach to address big data problems. Thirdly, WM-FRBS has been used to develop many kinds of big data classifiers because of its high-efficiency in modeling and inference procedures [48,49]. By comparing with CBRBS and WM-FRBS, it can be found that both of them need to generate rules from data samples with consideration of rule reduction and activate rules to produce inference outputs, resulting in the similar computing complexity.

$$\beta_n=\frac{(w_{j_1\cdots j_M}\beta_n^{j_1\cdots j_M}+1-w_{j_1\cdots j_M}\sum_{i=1}^{N}\beta_i^{j_1\cdots j_M})-(1-w_{j_1\cdots j_M}\sum_{i=1}^{N}\beta_i^{j_1\cdots j_M})}{\sum_{i=1}^{N}(w_{j_1\cdots j_M}\beta_n^{j_1\cdots j_M}+1-w_{j_1\cdots j_M}\sum_{i=1}^{N}\beta_i^{j_1\cdots j_M})-(N-1)(1-w_{j_1\cdots j_M}\sum_{i=1}^{N}\beta_i^{j_1\cdots j_M})-(1-w_{j_1\cdots j_M})}=\frac{\beta_n^{j_1\cdots j_M}}{\sum_{i=1}^{N}\beta_n^{j_1\cdots j_M}+(1-\sum_{i=1}^{N}\beta_i^{j_1\cdots j_M})}=\beta_n^{j_1\cdots j_M} \tag{34}$$

Box II.

Lemma 2 (*Boundedness Property*)**.** *The inference output of a CBRBS is bounded.*

Proof. According to the analytical ER algorithm in Eq. (7), it must be the case that the integrated belief degree meets the following constraints:

$$0\le\beta_n\le 1;\, n=1,\ldots,N \tag{30}$$

$$0\le\sum_{n=1}^{N}\beta_n\le 1 \tag{31}$$

Meanwhile, the inference output of the CBRBS has the following derivation processes:

$$\begin{aligned} f(\boldsymbol{x}) &= \sum_{n=1}^{N}u(D_n)\beta_n+(1-\sum_{n=1}^{N}\beta_n)\frac{u(D_1)+u(D_N)}{2}\\ &= \frac{u(D_1)+u(D_N)}{2}+\sum_{n=1}^{N}\frac{u(D_n)-u(D_1)-u(D_N)}{2}\beta_n \end{aligned} \tag{32}$$

Thus, due to $u(D_1)\le u(D_2)\le\ldots\le u(D_N)$, the inference output of the CBRBS will reach the lower bound when $\beta_1=1$ and $\beta_n=0$ ($n=2,\ldots,N$) and it will reach the upper bound when $\beta_n=0$ ($n=1,\ldots,N$), namely

$$\begin{aligned} \frac{u(D_1)}{2}\le f(\boldsymbol{x}) &= \frac{u(D_1)+u(D_N)}{2}+\sum_{n=1}^{N}\frac{u(D_n)-u(D_1)-u(D_N)}{2}\beta_n\\ &\le \frac{u(D_1)+u(D_N)}{2} \end{aligned} \tag{33}$$

The above results demonstrate that the inference output of the CBRBS is bounded.

Lemma 3 (*Continuity Property*)**.** *The inference output of the CBRBS is continuous.*

Proof. For a CBRBS, there are usually two kinds of situations of activating cumulative belief rules according to the activated rules determination shown in Section 4.1. When the number of activated rules is greater than one, it is obvious that the inference output of CBRBS is continuous based on the analytical rule integration process given in Eqs. (25)–(27). As a result, it only needs to prove the continuity under the situation of activating one cumulative belief rule.

When there is one activated rules $R_{j_1\cdots j_M}$, the calculation of activation weight in Eq. (26) shows that $w_{j_1\cdots j_M}=1$. Afterwards, the integrated belief degree can be simplified as given in Box II .

The above results show that the inference output of CBRBS is independent of any cumulative belief rule adjacent to $R_{j_1\cdots j_M}$ when there is one activated rule. As a result, it is proven that CBRBS is continuous.

According to Lemmas 1 to 3, it can be concluded that a CBRBS can produce an inference output for any given input data within linear computing complexity, and its inference output is bounded and continuous.

5.3. Universal approximation property of the CBRBS

In the previous studies [18,22,23], Stone–Weierstrass theorem was used to demonstrate the universal approximation of BRBS, WM-FRBS, and EBRBS. Hence, in this section, Stone–Weierstrass theorem can be also applied to illustrate the universal approximation of the CBRBS below.

Theorem 1 (*Universal Approximation Theorem*)**.** *For any given real continuous function $g(x)$, on a compact domain $U\subseteq R^m$ and arbitrary positive number $\varepsilon>0$, there exists a CBRBS $f(x)\in F(x)$ with $F(x)$ being the set of all the CBRBSs, such that*

$$\sup_{x\in U}|g(x)-f(x)|<\varepsilon \tag{35}$$

Based on the proof of the universal approximation property of the BRBS (Please see [22] for details), Theorem 1 will be hold as long as the CBRBS satisfies the conditions: (1) $f(x)$ ($f(x)\in F(x)$) is continuous; (2) $F(x)$ is an algebra, *i.e.*, $f_1(x)+f_2(x)\in F(x)$, $f_1(x)f_2(x)\in F(x)$, and $cf_2(x)\in F(x)$ ($f_1(x)$, $f_2(x)\in F(x)$); (3) $F(x)$ is separate points on U, *i.e.*, for every x, $x'\in U$, $x\neq x'$, there exists $f(x)\in F(x)$ such that $f(x)\neq f(x')$; (4) $F(x)$ vanishes at no point of U, *i.e.*, for each $x\in U$, there exists $f(x)\in F(x)$ such that $f(x)\neq 0$.

Apart from the condition (1) that has been proven in Lemma 3, the remaining conditions are proven in forms of three lemmas.

Lemma 4 (*Algebra Property*)**.** *The set of all the CBRBS $F(x)$ is an algebra, i.e, $f_1(x)+f_2(x)\in F(x)$, $f_1(x)f_2(x)\in F(x)$, and $cf(x)\in F(x)(f(x), f_1(x), f_2(x)\in F(x))$.*

Proof. For any input data vector $\boldsymbol{x}$, let CBRBS $f_1(x)$, $f_2(x)\in F(x)$, so that each of two models can be wrote as:

$$f_s(x)=\frac{\sum_{n_s=1}^{N_s}u^s(D_{n_s})\eta_{n_s}^s(x)}{\sum_{n_s=1}^{N_s}\eta_{n_s}^s(x)},\ s=1,2 \tag{36}$$

where

$$\eta_{n_s}^s(x)=\prod_{k=1}^{L'}(w_k\beta_{n_s}^k+1-w_k)-\prod_{k=1}^{L'}(1-w_k) \tag{37}$$

The above results show that the inference output of the CBRBS is equals to the inference output of the BRBS. Hence, based on the proof that the set of all BRBSs is an algebra in [22], we can prove that $f_1(x)+f_2(x)\in F(x)$, $f_1(x)f_2(x)\in F(x)$ and $cf(x)\in F(x)$. In summary, the set of all the CBRBS $F(x)$ is an algebra.

Lemma 5 (*Separate Point Property*)**.** *The set of all the CEBRBSs $F(x)$ is separate points on U, i.e, for every input data x, $x'\in U$ ($x\neq x'$), there exists $f(x)\in F(x)$ such that $f(x)\neq f(x')$.*

Proof. Without loss of generality, assume that a CBRB has two cumulative belief rules R_1 and R_2 with different belief distributions. Meanwhile, for every input data x, $x'\in U$ ($x\neq x'$), while the rule R_1 is the only activated rule for the input data x and the rule R_2 is the only activated rule for the input data x', it is easy to prove $f(x)\neq f(x')$ based on the rules R_1 and R_2 which have different distribution.

Lemma 6 (*Vanish Property*)**.** *The set of all the CBRBSs $F(x)$ vanishes at no point of U, i.e, for each $x\in U$, there exists $f(x)\in F(x)$ such that $f(x)\neq 0$.*

Table 7
Two extended belief rules for pipeline leak detection.

Rule No.	Rule weight	Flow difference (U_1)				Pressure difference (U_2)			Leak size (D)		
		$NL(A_{1,1})$	$NS(A_{1,2})$	$Z(A_{1,2})$	$PL(A_{1,3})$	$NL(A_{2,1})$	$Z(A_{2,2})$	$PL(A_{2,3})$	$Zero(D_1)$	$Middle(D_2)$	$Large(D_3)$
R_k	1	0	0	0.6	0.4	0	0.7	0.3	0	0.75	0.25
R_t	1	0	0.2	0.8	0	0.1	0.9	0	0.2	0.7	0

Proof. By analyzing the similarity measure shown in Section 5.2, we can activate one cumulative belief rule which has a non-zero similarity between rule antecedent and input data, and the resulting activation weight of this rule is greater than zero according to Eq. (26). Therefore, the integrated belief degrees $\beta_n(n = 1,\ldots, N)$ would be greater than zero based on Eq. (27). Furthermore, without loss of generality, while we assume that $u(D_n)^{\min} = \min_{n=1,\ldots,N}\{u(D_n)\} > 0$, the inference output of the CBRBS would be greater than zero for each $x \in U$.

Therefore, by using Stone–Weierstrass theorem together with Lemmas 3 to 6, we prove that CBRBS processes the universal approximation capability.

6. Experimental analysis based on regression and classification problems

In order to validate the performance of CBRBS, the explainability of CBRBS is firstly discussed in the case of pipeline leak detection. Two cases regarding regression and classification problems are studies to further compare the accuracy and high-efficiency of CBRBS with other existing methods.

6.1. Explainability in the CBRBS based on pipeline leak detection

In this subsection, the problem of pipeline leak detection is used to illustrate the explainability of CBRBS. Suppose that two attributes U_i ($i = 1, 2$), namely flow difference and pressure difference, have 4 and 3 referential values $A_{i,j}$ ($j = 1,\ldots, J_i$; $J_1 = 4$; $J_2 = 3$). These referential values and their utility values are ($A_{1,j}$, $u(A_{1,j})$)$\in${(NL, -10), (NS, -6), (Z, 0), (PL, 3)} and ($A_{2,j}$, $u(A_{2,j})$)$\in${(NL, -0.02), (Z, 0), (PL, 0.02)}, respectively. Consequent attribute D, namely leak size, has 3 consequents D_n ($n = 1, 2, 3$). These consequents and their utility values are (D_n, $u(D_n)$)$\in${($Zero$, 0), ($Middle$, 4), ($Large$, 8)}.

Based on the given utility values, when a historical input–output data pair of pipeline leak detection is $< \boldsymbol{x_k}, y_k> = < x_{k,1} = 1.2, x_{k,2} = 0.006, y_k = 5>$, one extended belief rule, denoted as R_k, can be generated by using adjacent utility-based function to transform input and output data into belief distributions. When there is lack of historical data, extended belief rules can be given directly by experts, *i.e.*, three belief distributions {(NL, 0.2), (Z, 0.8), (PL, 0)}, {(NL, 0.1), (Z, 0.9), (PL,0)} and {($Zero$, 0.2), ($Middle$, 0.7), ($Large$, 0)} are provided by experts to express another extended belief rule, denoted as R_t. Table 7 shows the belief distributions and rule weights of R_k and R_t.

From Table 7, it is clear that R_k and R_t have the maximum belief degree in the same referential value, namely the 2nd referential value Z of flow difference and the 2nd referential value Z of pressure difference, indicating that R_k and R_t can be identified as similar extended belief rules. Hence, a cumulative belief rule $R_{2\cdot 2}$ should be generated by using ER algorithm to integrate R_k and R_t, in which the cumulative belief rule is shown in Fig. 12.

Looking at Fig. 12, cumulative belief rule can provide an easy-to-understand explanation for knowledge representation, *i.e.*, the cumulative belief rule shown in Fig. 12 is described as: when 8.06% sure that flow difference is NS, 75.81% is Z and 16.13% is PL, and 3.8% sure that pressure difference is NL, 84.79% is Z and 11.41% is PL, then 7.62% sure that leak size is $Zero$, 78.10% is $Middle$ and 10.48% is $Large$. The total belief degree of leak size is 7.62% + 78.10% + 10.48% = 96.20% < 100%, thus this cumulative belief rule contains incomplete uncertainty because the total belief degree of leak size in R_t is 90%. Owing the above explainability, it helps decision-makers to fully understand the knowledge in CBRBS.

Furthermore, when 500 historical input–output data pairs collected from pipeline leak detection are used to generate cumulative belief rules, a total of five cumulative belief rules can be generated after transforming these 500 data pairs into 500 extended belief rules, in which these data and rules are shown in Table 8 and Fig. 13.

From Fig. 13, each historical data pair is used to generate an extended belief rule and further generate cumulative belief rules, resulting in a total of 500 extended belief rules and 5 cumulative belief rules. Due to the fact that the large number of rules in a rule-based system will clearly weaken interpretability, CBRBS provides an efficient rule-base modeling procedure to ensure its interpretability by integrating similar extended belief rules to downsize rule-base.

In order to give a full view of the explainability in CBRB inference procedure, another two data regarding pipeline with leak status and no-leak status are provided in the form of belief distributions, as shown in Table 9.

According to the two data shown in Table 9, the cumulative belief rules shown in Table 8 can be used to produce two corresponding inference outputs. The detailed process includes the use of the nearest neighbor strategy to activate consistent cumulative belief rules and the calculation of similarity for each antecedent attribute and activation weight for each activated rule. Table 10 shows activated rules and calculation results of the data regarding pipeline with leak status and no-leak status. From Table 10, three cumulative belief rules $R_{3\cdot 2}$, $R_{3\cdot 3}$ and $R_{2\cdot 2}$ should be activated to reply the data regarding pipeline with leak status because they have the better activation priority than the other rules. Meanwhile, the importance of $R_{2\cdot 2}$ is higher than that of $R_{3\cdot 2}$ and $R_{3\cdot 3}$ according to their activation weights. In the perspective of belief distribution for the consequent attribute of $R_{3\cdot 2}$, $R_{3\cdot 3}$ and $R_{2\cdot 2}$, it can be found that $R_{2\cdot 2}$ has bigger belief degrees in $Middle$ and $Large$ than $Zero$, but $R_{3\cdot 2}$ and $R_{3\cdot 3}$ are just the opposite. Correspondingly, cumulative belief rule $R_{3\cdot 2}$ is the only rule having the best activation priority and thus there is only one activated rule to reply the data regarding pipeline with no-leak status, in which $R_{3\cdot 2}$ has the biggest belief degree in $Zero$. There is suggestive evidence that CBRBS can provide the details of which cumulative belief rule to be activated and its importance in each decision-making process.

Furthermore, the belief distribution in consequent attribute and the activation weight of all activated rules are used to produce the inferential results of CBRBS for predicting the leak size of pipeline, in which the obtained belief distributions for the pipeline with leak status and no-leak status are shown in Fig. 14.

From Fig. 14, the inferential result can be explained as: 96.95% sure that leak size is $Zero$, 2.9% is $Middle$, and 0.15% is $Large$ when CBRBS is used to reply the data of pipeline with no-leak status, and 14.5% sure that leak size is $Zero$, 41.54% is $Middle$, and 43.96% is $Large$ for replying the data of pipeline with leak status. Moreover, when the utility value of each consequent is considered, namely $u(Zero) = 0$, $u(Middle) = 4$, and

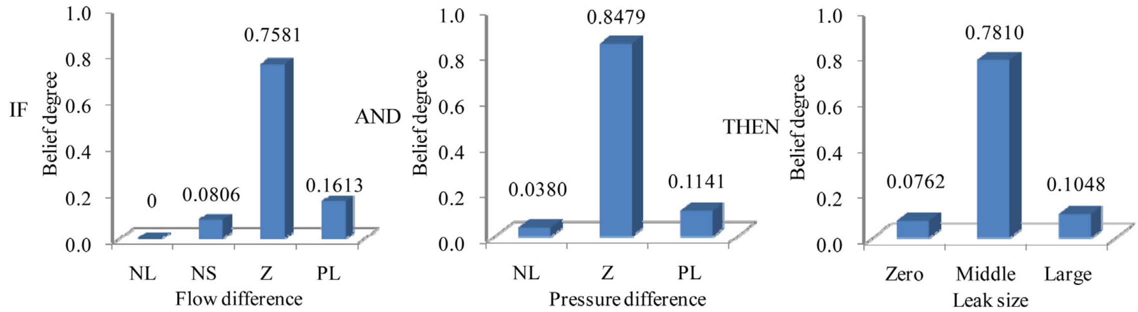


Fig. 12. One cumulative belief rule for pipeline leak detection.

Table 8
Five cumulative belief rules for pipeline leak detection.[1]

Rule No.	Rule weight	Flow difference (U_1)				Pressure difference (U_2)			Leak size (D)		
		$NL(A_{1,1})$	$NS(A_{1,2})$	$Z(A_{1,3})$	$PL(A_{1,4})$	$NL(A_{2,1})$	$Z(A_{2,2})$	$PL(A_{2,3})$	$Zero(D_1)$	$Middle(D_2)$	$Large(D_3)$
$R_{3\cdot2}$	0.758	0.0000	0.0274	<u>0.9477</u>	0.0249	0.0185	<u>0.9617</u>	0.0198	0.9695	0.0290	0.0015
$R_{3\cdot3}$	0.006	0.0000	0.0198	<u>0.9802</u>	0.0000	0.0000	0.4590	<u>0.5410</u>	1.0000	0.0000	0.0000
$R_{3\cdot1}$	0.004	0.0000	0.0000	<u>0.9070</u>	0.0930	<u>0.5449</u>	0.4551	0.0000	1.0000	0.0000	0.0000
$R_{2\cdot2}$	0.230	0.0316	<u>0.9557</u>	0.0127	0.0000	0.0165	<u>0.9682</u>	0.0154	0.0205	0.4714	0.5081
$R_{2\cdot1}$	0.002	0.1125	<u>0.8875</u>	0.0000	0.0000	<u>0.6501</u>	0.3500	0.0000	0.0000	0.4309	0.5691

Table 9
Two data in the form of belief distributions for pipeline leak detection.

Pipeline status	Leak size	Flow difference (U_1)				Pressure difference (U_2)		
		$NL(A_{1,1})$	$NS(A_{1,2})$	$Z(A_{1,2})$	$PL(A_{1,3})$	$NL(A_{2,1})$	$Z(A_{2,2})$	$PL(A_{2,3})$
Leak	6.3195	0.1375	0.8625	0	0	0	0.4500	0.5500
No-leak	0	0	0	0.9833	0.0167	0	1	0

Table 10
Rule activation explanation for pipeline leak detection.

Rule No.	Pipeline with leak status					Pipeline with no-leak status				
	$p_{j_1\cdot j_2}$	Activated	$Sim(U_1, x_1, R_{j_1\cdot j_2})$	$Sim(U_2, x_2, R_{j_1\cdot j_2})$	$w_{j_1\cdot j_2}$	$p_{j_1\cdot j_2}$	Activated	$Sim(U_1, x_1, R_{j_1\cdot j_2})$	$Sim(U_2, x_2, R_{j_1\cdot j_2})$	$w_{j_1\cdot j_2}$
$R_{3\cdot2}$	1	Yes	0.1014	0.4788	0.2720	1	Yes	0.9677	0.9668	1.0000
$R_{3\cdot3}$	1	Yes	0.0808	0.9910	0.0036	2	No	–	–	–
$R_{3\cdot1}$	2	No	–	–	–	2	No	–	–	–
$R_{2\cdot2}$	1	Yes	0.8998	0.4734	0.7244	2	No	–	–	–
$R_{2\cdot1}$	2	No	–	–	–	2	No	–	–	–

$u(Large) = 8$, the inferential result of CBRBS can be further transformed into a numeric output, *e.g.*, the predicted leak size is $0 \times 0.9695+4 \times 0.029+8 \times 0.0015 = 0.1281$ for the data of pipeline with no-leak status, whose real leak size is 0, and $0 \times 0.1450+4 \times 0.4154+8 \times 0.4396 = 5.1782$ for the data of pipeline with leak status, whose real leak size is 6.3195.

6.2. Comparative experiment based on pipeline leak detection

In this section, the problem of pipeline leak detection is re-examined to illustrate the prediction performance of BRBS, WM-FRBS, EBRBS, and CBRBS via two different comparative experiments. According to the previous studies [28,30,50], flow and pressure difference are assumed to have 8 and 7 referential values, respectively, leak size has 5 consequents. The utility values of referential values are shown in Eq. (16) and the utility values of consequents are {0, 2, 4, 6, 8}. 500 input–output data are randomly selected from 2008 data as training data and the 2008 data as testing data.

An Intelligent DEcision-mAking Helper (IdeaHelper[2]) has been developed to implement CBRBS and it empowers users to assess the behavior of modeling and inference using CBRBS for different kinds of decision-making problems, including regression problems and classification problems.

In the first comparative experiment, four kinds of EBRBSs, namely EBRBS, EBRBS with CBRB modeling procedure, EBRBS with CBRB inference procedure, and CBRBS are constructed using the two kinds of input transformation functions, namely adjacent utility-based and minimax utility-based functions, and the three kinds of similarity measure functions, namely SED-based, CED-based, and WED-based functions. Tables 11 to 14 show their results measured with (1) mean absolute error (MAE), which is a commonly used indicator to show the accuracy of prediction models when handling the pipeline leak detection problem [35, 51–53]; (2) number of rules; (3) number of failed data (number of data where the EBRBS could not activate any rules to produce an inference output); (4) rule activation ratio (average percentage of rules activated for each input data); (5) modeling time (the time of generating an EBRB from 500 training data in milliseconds); (6) inference time (the time of replying 2008 testing data in milliseconds).

From Tables 11 to 14, some conclusions can be summarized as follows:

(1) By comparing with three kinds of similarity measure functions in Tables 11 to 14, it can be found that the CED-based function fails to ensure that the model can activate one rule at least for replying given input data under the adjacent utility-based function. This is because the CED-based function has a

1 The maximum belief degree in two antecedent attributes is highlighted in underline.

2 IdeaHelper is available from the author via e-mail: more026@hotmail.com

Table 11
Comparison of EBRBSs with different system settings.

Indicator	Adjacent utility-based function			Minimax utility-based function		
	SED	CED	WED	SED	CED	WED
MAE	0.9256	0.2113	1.1509	1.6647	1.5385	1.6895
No. of rules	500	500	500	500	500	500
No. of failed data	0	2	0	0	0	0
Rule activation ratio	100	51.11	100	100	99.91	100
Modeling time (ms)	521	450	610	630	676	610
Inference time (ms)	1348	1234	1357	1819	1792	1674

Table 12
Comparison of the EBRBSs improved by CBRB modeling procedure.

Indicator	Adjacent utility-based function			Minimax utility-based function		
	SED	CED	WED	SED	CED	WED
MAE	0.7901	0.2089	1.0211	1.5735	1.4348	1.6002
No. of rules	21	21	21	24	24	24
No. of failed data	0	3	0	0	0	0
Rule activation ratio	100	23.02	100	100	99.90	100
Modeling time (ms)	39	24	24	25	39	19
Inference time (ms)	100	82	70	143	127	138

Table 13
Comparison of EBRBSs improved by CBRB inference procedure.

Indicator	Adjacent utility-based function			Minimax utility-based function		
	SED	CED	WED	SED	CED	WED
MAE	0.2034	0.1931	0.2047	0.2059	0.2055	0.2060
No. of rules	500	500	500	500	500	500
No. of failed data	0	2	0	0	0	0
Rule activation ratio	30.23	30.22	30.24	21.65	21.65	21.65
Modeling time (ms)	484	440	471	625	669	645
Inference time (ms)	403	423	378	398	395	370

Table 14
Comparison of CBRBS with different system settings.

Indicator	Adjacent utility-based function			Minimax utility-based function		
	SED	CED	WED	SED	CED	WED
MAE	0.2153	0.2089	0.2154	0.2069	0.2069	0.2069
No. of rules	21	21	21	24	24	24
No. of failed data	0	3	0	0	0	0
Rule activation ratio	4.80	4.77	4.80	4.20	4.20	4.20
Modeling time (ms)	19	18	15	24	26	18
Inference time (ms)	17	16	14	20	22	26

constraint condition to keep the Euclidean distance between two belief distributions less than 1. Otherwise, the similarity obtained by the CED-based function will be zero. Comparatively, the minimax utility-based function can help the CED-based similarity measure function overcome the issue that fails to activate any rule, so the number of failed data is zero.

(2) It is clear from Table 11 that, apart from the EBRBS with CED-based similarity measure function, all kinds of other EBRBSs have 500 rules and almost 100% rule activation ratio, which mean that all training data are generated into extended belief rules and all these rules have to activated for replying any given input data, namely Challenges 1 and 2 detailed in Section 2.2, so the resulting EBRBSs usually need a lot of time to perform the EBRB modeling and inference procedures under the adjacent utility-based and minimax utility-based functions, but their MAEs are all greater than 1.0.

(3) It is clear from Table 12 that the CBRB modeling procedure can decrease MAE, number of rules, modeling time, and inference time under the adjacent utility-based and minimax utility-based functions. This is because redundant rules are removed from the EBRB so that the reduced EBRBS can accurately predict the size of pipeline leak based on less number of rules. However, the rule activation ratios (exclude the CED-based function) are still near 100%, which mean that the EBRBS has to depend on conflicting information to reply new input data, namely Challenge 2 discussed in Section 2.2.

(4) It is clear from Table 13 that the CBRB inference procedure can decrease the total time of generating an EBRB and replying all input data, especially for the decrease of MAE and rule activation ratio. This is because the CBRB inference procedure can contribute to activate the most consistent rules, instead of all rules, for replying any give input data. However, the number of rules is still 500 for all kinds of EBRBs, which means that the size of EBRB is equal to the total number of training data, namely Challenge 1 shown in Section 2.2.

(5) It is clear from Table 14 that when an EBRBS is improved by using the CBRB modeling and inference procedures, a CBRBS can be constructed by upgrading an EBRBS and it has less number of rules and lower rule activation ratio than the other EBRBSs detailed in Tables 11 to 13. Meanwhile, the MAE, modeling time, and inference time of the CBRBS are also better than those of the EBRBSs shown in Tables 11 to 13. Thus, it can be concluded

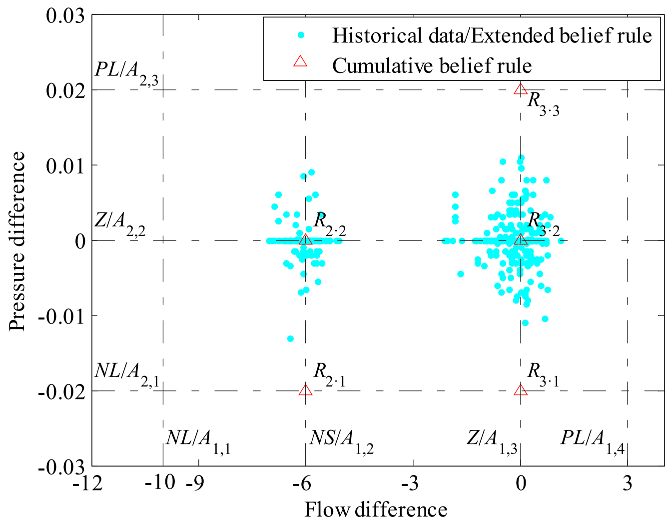


Fig. 13. Data and rule distributions of pipeline leak detection.

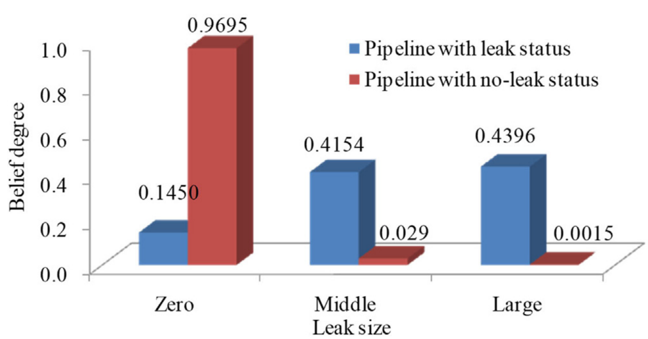


Fig. 14. Inferential results of CBRBS for pipeline leak detection.

that the CBRBS is able to overcome Challenges 1 and 2 shown in Section 2.2.

In the second comparative experiment, the BRBS and WM-FRBS are used to compare with the CBRBS under the two kinds of transformation functions and the three kinds of similarity measure functions. Table 15 shows their results measured with MAE, number of rules, number of failed data, rule activation ratio, modeling time, and inference time, in which the parameters used in the BRBS are determined by experts [28], instead of parameter learning. It is clear from Table 15 that the CBRBS has a lower MAE than the BRBS and WM-FRBS, especially for the minimax utility-based function which makes the BRBS and WM-FRBS have to activate all rules for replying any given input data (More details can be found in Section 4.3). In the comparison of BRBS and CBRBS, it can be found that the MAE of BRBS is significantly worse than CBRBS but the generation of a BRB is 0 ms. This is because the initial BRB is generated by using domain knowledge without any calculation process, but this initial BRB sometime fails to ensure a desired prediction performance of the BRBS due to limited domain knowledge.

For the purpose of further comparing BRBS and CBRBS, the parameter learning model shown in Sections 2.4 and 5.1 are used to optimize the parameters of BRBS and CBRBS, respectively, in which the differential evolution (DE) algorithm [35] is introduced to solve the parameter learning model under the termination criteria that the fitness of the best parameter vector does not change during the last 100 iterations. Table 16 shows the results of trained EBRBS and CBRBS under the tradition utility-based and minimax utility-based functions. From Table 16, it is clear that the MAE of trained BRBS can be significantly decreased from 1.0384 to 0.2707 after 1180 iterations when using the adjacent utility-based function but failed under the minimax utility-based function. This is because the BRB inference has to activate all rules for any given input data when the belief distribution generated from the input data has non-zero belief degrees in any referential value. By comparing the CBRBS shown in Table 15 and the trained CBRBS in Table 16, it can be found that the parameter learning model can slightly improve the prediction performance of CBRBS, *i.e.*, the MAE of CBRBS decreases from 0.2154 to 0.2022, but the cost of this improvement is the need of a lot of training time. In addition, it is worth noting that the BRBS's number of the parameters needed to train is greater than CBRBS owing to the fact that almost all kinds of parameters in a BRB, *e.g.*, belief degrees, utility values, rule weights, and attribute weights, should be optimized using the parameter learning model. But for CBRBS, the parameters needed to train only include utility values and attribute weights.

In order to further compare trained BRBSs with CBRBS, ten existing studies on the BRBS with parameter learning are introduced and their results can be found in Table 17. By comparing the results of selected studies and CBRBS, it is validated that the CBRBS can produce a satisfactory result on the pipeline leak detection problem. More importantly, the result of CBRBS does not depend on parameter learning, so CBRBS has an efficient rule-base modeling procedure.

In summary, for the comparison of BRBS, WM-FRBS, EBRBS, and CBRBS on pipeline leak detection problem, the experiment results have shown that CBRBS can effectively and efficiently produce inference outputs to replying input data, whose prediction performance is better than BRBS, WM-FRBS, and EBRBS.

6.3. Comparative experiment based on classification problems

In this section, 16 classification datasets obtained from the well-known UCI machine-learning repository are used to validate the classification performance of CBRBS through three different comparative experiments. The main characteristics of these 18 classification datasets are summarized in Table 18, where "#Data" denotes the number of data, "#Attribute" denotes the number of attributes, and "#Class" denotes the number of classes.

To construct CBRBS for each classification dataset, suppose that each antecedent attribute of each dataset has seven referential values and the number of consequents is equal to the number of classes. In addition, 10-fold cross-validation (10-CV) is applied to test each dataset. In other words, each classification dataset is divided into 10 blocks, with 9 blocks as a training dataset and the remaining block as a testing dataset. Noting that the classification accuracy, which is the most common indicator to evaluate the performance of classifiers [24,39,48,57], is used to measure the performance of CBRB and it is obtained based on two kinds of transformation functions and three kinds of similarity measure functions. Three different comparative experiments are provided as follows:

In the first comparative experiment, seven kinds of improved EBRBS in previous studies are introduced to compare with CBRBS. Table 19 shows the basic information and the accuracy of seven classification datasets for the EBRBS, seven improved EBRBSs, and CBRBS. Note that the results of all these EBRBSs are obtained from the EBRB without parameter learning because the parameter learning is a time-consuming process and will significantly weaken the high efficiency of the EBRB modeling.

It is clear from Table 19 that, on the one hand, although all of these EBRBSs proposed in previous studies aimed to solve

Table 15
Comparison of BRBS, WM-FRBS, and CBRBS.

Indicator	Adjacent utility-based function			Minimax utility-based function		
	BRBS	WM-FRBS	CBRBS	BRBS	WM-FRBS	CBRBS
MAE	1.0384	0.2274	0.2154	1.5274	1.3442	0.2069
No. of rules	56	21	21	56	24	24
No. of failed data	0	2	0	0	1	0
Rule activation ratio	4.51	18.33	4.80	100	99.74	4.20
Modeling time (ms)	0	29	17	0	18	17
Inference time (ms)	13	20	15	39	5	16

Table 16
Comparison of BRBS and CBRBS with parameter learning.

Indicator	Adjacent utility-based function		Minimax utility-based function	
	Trained BRBS	Trained CBRBS	Trained BRBS	Trained CBRBS
MAE	0.2707	0.2022	1.5274	0.2029
No. of rules	56	32	56	25
No. of failed data	0	0	0	0
Rule activation ratio (%)	7.14	3.16	100	4.06
No. of parameters to train	358	20	358	20
No. of iterations	1180	309	101	358
Training time (ms)	167442	334626	71215	407934

Table 17
Comparison of CBRBS with existing studies.

No.	Year	Description	Training	MAE	Size (testing data)	No. of parameters to train	No. of rules
1	2007 [46]	Local training	Yes	0.2223	2008	353	56
2	2009 [50]	Online updating	Yes	0.3954	2008	353	56
3	2010 [54]	Sequential learning	Yes	0.8506	17	34	5
4	2011 [53]	Adaptive training	Yes	0.2064	2008	353	56
5	2015 [52]	Approximate causal inference	Yes	0.2014	2008	92	15
6	2016 [55]	Dynamic rule adjustment	Yes	0.2080	2008	43	6
7	2018 [51]	Bi-level BRB	Yes	0.1941	2008	40	5
8	2018 [56]	JOPS algorithm	Yes	0.1738	2008	43	6
9	2018 [57]	Akaike information criterion	Yes	0.2339	2008	22	8
10	2019 [58]	Online updating	Yes	0.3900	800	41	6
11	2021 [59]	SGDM-BRB	Yes	0.2024	2008	114	16
12	2021 [60]	PMP-BRB	Yes	0.3130	2008	40	5
11	This study	CBRBS	No	0.2069	2008	0	21

Table 18
Basic information of classification datasets.

No.	Dataset	#Data	#Attribute	#Class	No.	Dataset	#Data	#Attribute	#Class
1	Appendicitis	106	7	2	10	Magic	19,020	10	2
2	Heart	270	13	2	11	Iris	150	4	3
3	Diabetes	393	8	2	12	Wine	178	13	3
4	Wdbc	569	30	2	13	Cleveland	297	13	5
5	Transfusion	748	4	2	14	Glass	214	9	6
6	Pima	768	8	2	15	Satimage	6,435	36	6
7	Mammographic	830	5	2	16	Penbased	10,992	16	10
8	Phoneme	5,404	5	2	17	Vowel	990	13	11
9	Twonorm	7,400	20	2	18	Texture	5,500	40	11

Challenges 1 and 2, all of them only focus on Challenges 1 or 2. Only the present study is the first time to solve both two challenges together by proposing generic rule-base modeling and inference procedures for upgrading the EBRB as the CBRBS. On the other hand, by comparing the classification accuracy of all these EBRBSs and CBRBS, it can be found that there is no universally applicable EBRBS that can reach the best accuracy for all classification datasets, *i.e.*, the CBRBS obtains the best accuracy on datasets Diabetes and Transfusion, the SO-EBRBS obtains the best accuracy on datasets Mammographic and Iris, the DBSCAN-EBRBS obtains the best accuracy on dataset Wine, and the CABRA-EBRBS obtains the best accuracy on datasets Wdbc and Glass. From the view of average rank, it is proved that the CBRBS outperform all list studied shown in Table 19. In the second comparative experiment, five kinds of classical FRBS-related classifiers are introduced to compare with CBRBS and they are structural learning algorithm on vague environment (SLAVE) [16], fuzzy hybrid generic-based machine learning algorithm (FH-GBML) [62], steady-state genetic algorithm for extracting fuzzy classification rules from data (SGRED) [29], fuzzy

Table 19
Comparison of CBRBS with existing EBRBSs.

Indicator/Dataset	EBRBS	Improved EBRBS									CBRBS
		DRA	MaSF	DEA	CABRA	DBSCAN	SO	BT	Micro	CT	
Year	2013 [23]	2015 [4]	2016 [38]	2017 [30]	2018 [41]	2020 [34]	2020 [43]	2020 [40]	2021 [24]	2021 [61]	This study
Solve Challenge 1	No	No	No	Yes	No	Yes	No	No	Yes	No	Yes
Solve Challenge 2	No	Yes	Yes	No	Yes	No	Yes	Yes	No	Yes	Yes
Wdbc	94.59 (9)	94.61 (8)	96.30 (4)	–	97.01 (1)	95.47 (6)	95.22 (7)	96.00 (5)	96.49(2.5)	–	96.49 (2.5)
Diabetes	73.39 (7)	71.44 (9)	75.04 (4.5)	72.77 (8)	76.34 (2)	75.98 (3)	75.04 (4.5)	–	74.91 (6)	–	76.84 (1)
Mammographic	77.64 (8)	78.39 (7)	80.61 (3)	–	79.52(6)	79.57 (5)	83.95 (1)	79.80 (4)	–	–	81.08 (2)
Iris	95.20 (8.5)	95.50 (7)	95.20 (8.5)	96.00 (3)	96.00 (3)	95.73 (4)	97.67 (1)	95.99 (5)	–	95.93 (6)	96.00 (3)
Wine	96.32 (7)	96.46 (6)	96.52 (4)	–	96.63 (2.5)	97.87 (1)	96.49 (5)	94.76 (9)	95.84 (8)	–	96.63 (2.5)
Transfusion	76.14 (5)	76.57 (3)	78 .65 (2)	–	72.07 (6)	57.33 (7)	–	–	76.52 (4)	–	79.14 (1)
Glass	67.85 (7)	69.65 (5)	70 .19 (3)	68.18 (6)	72.90 (1)	57.07 (9)	–	70.01 (4)	63.32 (8)	68.30 (5)	71.03 (2)
Average rank	7.4	6.4	4.1	5.7	3.1	5.0	3.7	5.4	5.7	5.5	2.5

Table 20
Comparison of CBRBS with classical FRBS-related classifiers.

Dataset	SLAVE	FH-GBML	SGRED	FARC-HD	CFAR	CBRBS
Appendicitis	82.91 (6)	86.00 (2)	84.48 (4)	84.18 (5)	87.82 (1)	85.85 (3)
Heart	71.36 (6)	75.93 (4)	73.21 (5)	84.44 (1)	82.22 (2)	80.37 (3)
Wdbc	92.33 (3)	92.26 (4)	90.68 (5)	95.25 (2)	–	96.49 (1)
Pima	73.71 (4)	75.26 (2)	73.37 (5)	75.66 (1)	65.11 (6)	74.09 (3)
Phoneme	76.41 (4)	79.66 (3)	75.55 (5)	82.14 (2)	70.65 (6)	82.42 (1)
Twonorm	86.99 (4)	85.97 (5)	73.98 (6)	95.28 (2)	91.66 (3)	96.77 (1)
Magic	73.96 (4)	81.30 (2)	72.06 (5)	84.51 (1)	64.84 (6)	80.61 (3)
Iris	94.44 (4)	94.00 (5)	94.89 (3)	96.00 (1.5)	90.67 (6)	96.00 (1.5)
Wine	89.47 (6)	92.61 (4)	91.88 (5)	94.35 (2)	93.24 (3)	96.63 (1)
Cleveland	48.82 (6)	53.51 (4)	51.59 (5)	55.24 (2)	53.88 (3)	55.89 (1)
Satimage	81.69 (3)	74.72 (5)	77.10 (4)	87.32 (2)	–	90.46 (1)
Penbased	81.16 (3)	50.45 (5)	67.93 (4)	96.04 (2)	36.43 (5)	99.14 (1)
Vowel	71.11 (3)	67.07 (4)	65.83 (5)	71.82 (2)	–	98.38 (1)
Texture	81.57 (3)	70.15 (5)	71.66 (4)	92.89 (2)	–	98.13 (1)
Average rank	4.2	3.9	4.6	2.0	4.1	1.6

association rule-based classification method for high-dimensional problem (FARC-HD) [4], and classification with fuzzy association rules (CFAR) [63], in which SGRED is one kind of improved WM-FRBS used for classification problems. Table 20 shows the classification accuracy of fourteen classification datasets related to the above five FRBS-related classifiers and these results are all obtained from [64].

From Table 20, it is clear that the CBRBS can obtain the best accuracy on ten of fourteen datasets and the third best accuracy on the other datasets, whose number is greater than other five kinds of FRBS-related classifiers. Hence, in the comparison of average rank, the CBRBS outperforms five kinds of FRBS-related classifiers and the order of average rank is CBRBS (1.6) > FARC-HD (2.0) > FH-GBML (3.9) > CFAR (4.1) > SLAVE (4.2) > SGRED (4.6). Additionally, it is worth noting that, apart from SGRED, all other FRBS-related classifiers belong to evolutionary fuzzy systems, which indicate that evolutionary algorithms should be used to iteratively update the parameters and/or structure of FRBS, so they usually need a large amount of computing time to generate a rule-base.

In the third comparative experiment, five kinds of classical machine-learning (ML)-related classifiers are introduced to compare with the CBRBS and they are k nearest neighbor (KNN) [4], naïve Bayes (NB) [65], decision tree (DT) [66], support vector machine (SVM) [67], and artificial neural network (ANN) [68]. The results of all these ML-related classifiers are obtained from the open source software WEKA. Table 21 shows the classification accuracy of eighteen datasets.

From Table 21, although the accuracy of the CBRBS is even worse than ML-related classifiers on some datasets like Heart (The best classifier is NB), Phoneme (The best classifier is DT), Iris (The best classifier is SVM and ANN) and Wine (The best classifier is KNN), it is still possible to see a considerable result because the CBRBS obtains the best accuracy on seven of eighteen datasets, whose number is greater than other ML-related classifiers. Consequently, in the comparison of average rank, CBRBS outperforms the five kinds of ML-related classifiers and the order of average rank is CBRBS (2.2) > ANN (2.8) > DT (3.6) > NB (3.7) > SVM (4.2) > KNN (4.4).

In summary, for the comparison of improved EBRBSs, FRBS-related and ML-related classifiers on some classification problems, the experiment results have shown that CBRBS not only can overcome the two challenges of EBRBS, but also has desired classification accuracy better than existing improved EBRBSs. Moreover, the classification performance of CBRBS can outperform some FRBS-related and ML-related classifiers.

7. Conclusion and discussion

This study proposed a highly interpretable rule-based system, named CBRBS, with focus on establishing its efficient rule-base modeling and inference procedures. The main motivation is to address the challenges of traditional rule-based systems and also try to find a proper balance between accuracy, high-efficiency and explainability under the background of the white-box modeling of XAI techniques. Detailed comparison analyses were carried out to illustrate the difference among WM-FRBS, BRBS, EBRBS, and CBRBS. Comparative experiments were also provided to exhaustively demonstrate the advantages of CBRBS. The detailed conclusions can be summarized as follows:

Table 21
Comparison of CBRBS with classical ML-related classifiers.

Dataset	KNN	NB	DT	SVM	ANN	CBRBS
Appendicitis	82.08 (5)	85.85 (2.5)	85.85 (2.5)	80.19 (6)	85.85 (2.5)	85.85 (2.5)
Heart	74.81 (5)	83.70 (1)	77.41 (4)	55.56 (6)	82.22 (2)	80.37 (3)
Diabetes	74.09 (4)	76.30 (2)	73.82 (5)	65.10 (6)	75.39 (3)	76.84 (1)
Wdbc	95.96 (2)	92.97 (5)	93.32 (4)	62.74 (6)	96.31 (3)	96.49 (1)
Transfusion	76.20 (4)	75.40 (5)	78.34 (2)	75.27 (6)	76.34 (3)	79.14 (1)
Pima	68.83 (6)	77.27 (1)	71.82 (5)	75.45 (2)	73.77 (4)	74.09 (3)
Mammographic	79.04 (6)	82.41 (2)	83.98 (1)	79.88 (5)	80.60 (4)	81.08 (3)
Phoneme	79.44 (5)	76.05 (6)	86.42 (1)	84.49 (2)	80.98 (4)	82.42 (3)
Twonorm	94.86 (5)	97.83 (1)	85.12 (6)	97.69 (2)	96.96 (3)	96.77 (4)
Magic	74.79 (4)	72.69 (5)	81.17 (2)	65.88 (6)	83.73 (1)	80.61 (3)
Iris	94.67 (4.5)	94.00 (6)	94.67 (4.5)	97.33 (1.5)	97.33 (1.5)	96.00 (3)
Wine	97.19 (1)	96.63(3.5)	92.13 (4)	44.38 (5)	97.17 (2)	96.63 (3.5)
Cleveland	55.56 (3)	54.88 (4)	56.57 (1)	53.87 (5)	52.53 (6)	55.89 (2)
Glass	66.36 (5)	48.60 (6)	66.82 (4)	68.69 (2)	67.76 (3)	71.03 (1)
Satimage	86.67 (2)	75.39 (6)	82.34 (5)	82.71 (4)	85.61 (3)	90.46 (1)
Penbased	55.90 (5)	85.68 (3)	74.35 (4)	13.71 (6)	94.39 (2)	99.14 (1)
Vowel	15.25 (6)	67.07 (3)	46.97 (5)	88.48 (2)	84.14 (3)	98.38 (1)
Texture	72.64 (6)	77.42 (4)	73.47 (5)	96.47 (3)	99.82 (1)	98.13 (2)
Average rank	4.4	3.7	3.6	4.2	2.8	2.2

(1) A generic procedure of CBRB modeling, including transformation of inputs to belief distributions, generation of extended belief rules from historical input–output data pairs, and integration of extended belief rules to generate cumulative belief rules using ER algorithm, is proposed to ensure the reasonable size of a CBRB so that CBRBS is able to overcome the challenge of rule boundlessness.

(2) A generic procedure of CBRB inference, including activation of cumulative belief rules using nearest neighbor strategy and the integration of activated cumulative belief rules using ER algorithm, is proposed to select consistent rules as activated rules and then integrate the activated rules to produce an inference output, so that CBRBS is able to overcome the challenge of rule inconsistency.

(3) The inherent properties, including linear time complexity, boundedness, continuity properties, and universal approximation theorem, are provided and proved for CBRBS. Moreover, the explainability of CBRBS is discussed on the basis of the pipeline leak detection to demonstrate its white-box characteristics for decision-maker. These properties and the characteristics contribute to facilitate the applications of CBRBS.

For future researches, additional parameter learning, structure learning, and/or joint learning for optimizing parameter and structure of CBRB should be undertaken. Moreover, it is necessary and useful to hybridize CBRBS with some existing rule-based fuzzy systems for a more interpretable and powerful rule-based system in the aim of solving different regression and classification problem. The CBRBS needs to be validated on more open benchmarks and practical problems.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This research was supported by the National Natural Science Foundation of China (Nos. 72001043 and 61773123), the Humanities and Social Science Foundation of the Ministry of Education of China (No. 20YJC630188), the National Science Foundation of Fujian Province of China (No. 2020J05122), and the Chengdu International Science Cooperation Project (No. 2020-GH02-00064-HZ).

Appendix A. Formula derivation of assigning rule weight of each extended belief rule

According to [24], the time complexity of calculating rule weights for T extended belief rules with J_i $(i = 1,\ldots, M)$ referential values and N consequents is $O(T^2 \times (\sum_{i=1}^{M} J_i + N))$. Hence, a speediness strategy is investigated to assign rule weights for each extended belief rule.

Firstly, as shown in Eq. (5) at Section 2.2, rule weight θ_k $(k = 1,\ldots, T)$ of each extended belief rule R_k is calculated based on inconsistency degrees $Incons(R_k)$. Hence, when $Incons(R_k)$ is assumed to be a variable, the first order partial derivative of the rule weight with respect to the inconsistency degree can be obtained as follows:

$$\frac{\partial \theta_k}{\partial Incons(R_k)} = -\frac{\sum_{l=1}^{T} Incons(R_l) - Incons(R_k)}{\left[Incons(R_k) + \sum_{l=1,l\neq k}^{T} Incons(R_l)\right]^2} = -\frac{\sum_{l=1,l\neq k}^{T} Incons(R_l)}{\left[Incons(R_k) + \sum_{l=1,l\neq k}^{T} Incons(R_l)\right]^2} < 0 \quad (A.1)$$

From Eq. (A.1), it is clear that rule weight θ_k decreases when its inconsistency degree $Incons(R_k)$ increases. In other words, when inconsistency degree $Incons(R_k)$ is equal to 1, rule R_k has a minimum rule weight. This makes sense that if this extended belief rule causes contradiction, then it will be useless.

Afterwards, to ensure the minimum rule weight for each extended belief rule, the inconsistency degree of all extended belief rules is assumed as 1, namely $Incons(R_k)$=1. Therefore, when a large number of extended belief rules (assume T is approaching to ∞) need to calculate rule weights, the rule weight for rule R_k would be equivalent to the following equality:

$$\theta_k = \lim_{T\to+\infty}\left[1 - \frac{Incons(R_k)}{\sum_{j=1}^{T} Incons(R_j)}\right] = 1 - \lim_{T\to+\infty}\frac{1}{T} = 1 \quad (A.2)$$

Appendix B. Pseudocode of CBRB modeling procedure

Pseudocode to generate cumulative belief rules for CBRB

For each training data ($\boldsymbol{x}_t$, y_t) in {($\boldsymbol{x}_t$,y_t); t=1,…,T}

  For each input data $x_{t,i}$ in $\boldsymbol{x}_t$={$x_{t,1}$,…,$x_{t,M}$}

    To calculate belief degree $a^t_{i,j}$ based on Definition 1 or 2;

  End for

  To calculate belief degree β^t_n for output data y_t based on Definition 1 or 2;

  $R_t =< a^t_{i,j}, \beta^t_n; i=1,..., M; j=1,..., J_i; n=1,..., N>$;

  To update rule-base $EBRB$=$EBRB \cup$ {R_t};

End for

For each extended belief rule R_k in $EBRB$

  For each antecedent attribute U_i in {U_1,… U_M}

    To find the index of the maximum belief degree using $j_i = \arg\max_{j=1,...,J_i}\{\alpha^k_{i,j}\}$;

  End for

  To update rule set $\boldsymbol{RC}_{j_1\cdots j_M} = \boldsymbol{RC}_{j_1\cdots j_M} \cup R_k$;

End for

For each rule set $\boldsymbol{RC}_{j_1\cdots j_M}$ in { $\boldsymbol{RC}_{j_1\cdots j_M}$; j_i=1,…, J_i; i=1,…, M}

  For each antecedent attribute U_i in {U_1,… U_M}

    To calculate belief degree $\alpha^{j_1\cdots j_M}_{i,j}$ using the ER algorithm (As shown in Eq. (12)) ;

  End for

  To calculate belief degree $\beta^{j_1\cdots j_M}_n$ using the ER algorithm (As shown in Eq. (13)) ;

  To calculate rule weight $\theta_{j_1\cdots j_M}$ (As shown in Eq. (14)) ;

  $R_{j_1\cdots j_M} =< a^{j_1\cdots j_M}_{i,j}, \beta^{j_1\cdots j_M}_n, \theta_{j_1\cdots j_M}; j_i=1,..., J_i; i=1,..., M; n=1,..., N>$;

  To update rule-base $CBRB = CBRB \cup R_{j_1\cdots j_M}$;

End for

Appendix C. Pseudocode of CBRB inference procedure

Pseudocode to reply input data using CBRB

For each testing data $\boldsymbol{x}_t$ in {$\boldsymbol{x}_1$,…,$\boldsymbol{x}_S$}

  For each input data $x_{t,i}$ in $\boldsymbol{x}_t$={$x_{t,1}$,…,$x_{t,M}$}

    To calculate belief degree $a^t_{i,j}$ based on Definition 1 or 2;

  End for

  For each extended belief rule $R_{j_1\cdots j_M}$ in $CBRB$

    For each antecedent attribute U_i in {U_1,…,U_M}

      To calculate activation priority $p_{j_1\cdots j_M}$ based on Eq. (17)

    End for

    To update the set of activated rules $AR(\boldsymbol{x}_t)$ based on Eq. (19)

  End for

  For each extended belief rule $R_{j_1\cdots j_M}$ in $AR(\boldsymbol{x}_t)$

    For each antecedent attribute U_i in {U_1,…,U_M}

      To calculate similarity $Sim(U_i, x_{t,i}, R_{j_1\cdots j_M})$ based on Definition 3, 4, 5 or 6;

    End for

    To calculate rule weight $w_{j_1\cdots j_M}$ based on Eq. (25);

  End for

  To calculate integrated belief degree β_n based on Eq. (26);

  To produce an inference output based on Eq. (27) or (28);

End for

References

[1] H. Hagras, Toward human-understandable, explainable AI, Computer 51 (9) (2018) 28–36.
[2] D. Polap, M. Wlodarczyk-Sielicka, N. Wawrzyniak, Automatic ship classification for a riverside monitoring system using a cascade of artificial intelligence techniques including penalties and rewards, ISA Trans. (2021) http://dx.doi.org/10.1016/j.isatra.2021.04.003, (In Press).
[3] X. Bai, X. Wang, X. Liu, Q. Liu, J. Song, N. Sebe, B. Kim, Explainable deep learning for efficient and robust pattern recognition: A survey of recent developments, Pattern Recognit. 120 (2021) 108102.
[4] F. Aghaeipoor, M.M. Javidi, A. Ferbandez, IFC-BD: An interpretaable fuzzy classifier for boosting explainable artificial intelligence in big data, IEEE Trans. Fuzzy Syst. (2021) http://dx.doi.org/10.1109/TFUZZ.2021.3049911, (In Press).
[5] L.X. Wang, Fast training algorithm for deep convolutional fuzzy systems with application to stock index prediction, IEEE Trans. Fuzzy Syst. 28 (7) (2020) 1301–1314.
[6] J.M. Mendel, P.P. Bonissone, Critical thinking about explainable AI (XAI) for rule-based fuzzy systems, IEEE Trans. Fuzzy Syst. 29 (12) (2021) 3579–3593.
[7] S. Sun, Robust reasoning: Integrating rule-based and similarity-based reasoning, Artificial Intelligence 75 (2) (1995) 241–295.
[8] S. Sachan, J.B. Yang, D.L. Xu, Benavides.David-Eraso.Li. Yang, An explainable AI decision-support-system to automate loan underwriting, Expert Syst. Appl. 144 (2020) 113100.
[9] J. Furnkranz, T. Kliegr, A Brief Overview of Rule Learning. RuleML: Rule Technologies: Foundations, Tools, & Applications, Springer, Cham, 2015.
[10] A. Gudys, M. Sikora, L. Wróbel, RuleKit: A comprehensive suite for rule-based learning, Knowl.-Based Syst. 194 (2020) 105480.
[11] J. Furnkranz, . KliegrT, . PaulheimH, On cognitive preferences and the plausibility of rule-based models, Mach. Learn. 109 (2020) 853–898.
[12] A. Fernandez, F. Herrera, O. Cordon, M.J. del Jesus, F. Marcelloni, Evolutionary fuzzy systems for explainable artificial intelligence: why, when, what for, and where to? IEEE Comput. Intell. Mag. 14 (1) (2019) 69–81.
[13] E.H. Mamdani, S. Assilian, An experiment in linguistic synthesis with a fuzzy logic controller, Int. J. Man-Mach. Stud. 7 (1) (1975) 1–13.
[14] D. Aha, D. Kibler, Instance-based learning algorithm, Mach. Learn. 6 (1991) 37–66.
[15] O. Cordón, M.J. Del Jesus, F. Herrera, M. Lozano, MOGUL: A methodology to obtain genetic fuzzy rule-based systems under the iterative rule learning approach, Int. J. Intell. Syst. 14 (11) (1999) 1123–1153.
[16] A. González, R. Pérez, SLAVE: A genetic learning system based on an iterative approach, IEEE Trans. Fuzzy Syst. 7 (2) (1999) 176–191.
[17] V. Ojha, A. Abraham, V. Snášel, Heuristic design of fuzzy inference systems: A review of three decades of research, Eng. Appl. Artif. Intell. 85 (2019) 845–864.
[18] L.X. Wang, J.M. Mendel, Generating fuzzy rules by learning from examples, IEEE Trans. Syst. Man Cybern. 22 (6) (1992) 1414–1427.
[19] J.B. Yang, J. Liu, J. Wang, H.S. Sii, H.W. Wang, Belief rule-base inference methodology using the evidential reasoning approach - RIMER, IEEE Trans. Syst. Man Cybern. - Part A: Syst. Humans 36 (2) (2006) 266–285.
[20] L.L. Chang, J.B. Sun, J. Jiang, M.J. Li, Parameter learning for the belief rule base system in the residual life probability prediction of metalized file capacitor, Knowl.-Based Syst. 73 (2015) 69–80.
[21] L.L. Chang, Y. Zhou, J. Jiang, M.J. Li, X.H. Zhang, Structure learning for belief rule base expert system: A comparative study, Knowl.-Based Syst. 39 (2013) 159–172.
[22] Y.W. Chen, J.B. Yang, D.L. Xu, S.L. Yang, On the inference and approximation properties of belief rule based systems, Inform. Sci. 234 (2013) 121–135.
[23] J. Liu, L. Martinez, A. Calzada, H. Wang, A novel belief rule base representation, generation and its inference methodology, Knowl.-Based Syst. 53 (2013) 129–141.
[24] L.H. Yang, J. Liu, Y.M. Wang, L. Martinez, A micro-extended belief rule-based system for big data multi-class classification problems, IEEE Trans. Syst. Man Cybern. Syst. 51 (2021) 420–440.
[25] A.P. Dempster, Upper and lower probabilities induced by a multivalued mapping, Ann. Math. Stat. 38 (2) (1967) 325–339.
[26] G. Shafer, A Mathematical Theory of Evidence, Princeton University Press, Princeton, 1976.
[27] J.B. Yang, Rule and utility based evidential reasoning approach for multi-attribute decision analysis under uncertainties, European J. Oper. Res. 131 (1) (2001) 31–61.
[28] D.L. Xu, L. Liu, J.B. Yang, G.P. Liu, J. Wang, I. Jenkinson, J. Ren, Inference and learning methodology of belief- rule-based expert system for pipeline leak detection, Expert Syst. Appl. 32 (1) (2007) 103–113.
[29] E.G. Mansoori, M.J. Zolghadri, S.D. Katebi, SGERD: A steady-state genetic algorithm for extracting fuzzy classification rules from data, IEEE Trans. Fuzzy Syst. 16 (4) (2008) 1061–1071.
[30] L.H. Yang, Y.M. Wang, Y.X. Lan, L. Chen, Y.G. Fu, A data envelopment analysis (DEA)-based method for rule reduction in extended belief-rule-based systems, Knowl.-Based Syst. 123 (2017) 174–187.
[31] A. Calzada, J. Liu, H. Wang, A. Kashyap, A new dynamic rule activation method for extended belief rule-based systems, IEEE Trans. Knowl. Data Eng. 27 (4) (2015) 880–894.
[32] L. Jin, J. Liu, Y. Xu, X. Fang, A novel rule base representation and its inference method using the evidential reasoning approach, Knowl.-Based Syst. 87 (2015) 80–91.
[33] F.F. Ye, L.H. Yang, Y.M. Wang, L. Chen, An environmental pollution management method based on extended belief rule base and data envelopment analysis under interval uncertainty, Comput. Ind. Eng. 144 (2020) 106454.
[34] A. Zhang, F. Gao, M. Yang, W.H. Bi, A new rule reduction and training method for extended belief rule base based on DBSCAN algorithm, Internat. J. Approx. Reason. 119 (2020) 20–39.
[35] L.H. Yang, J. Liu, Y.M. Wang, L. Martínez, New activation weight calculation and parameter optimization for extended belief rule-based system based on sensitivity analysis, Knowl. Inf. Syst. 60 (2019) 837–878.
[36] M. Espinilla, J. Medina, A. Calzada, J. Liu, L. Martínez, C. Nugent, Optimizing the configuration of an heterogeneous architecture of sensors for activity recognition, using the extended belief rule-based inference methodology, Microprocess Microsyst. 52 (2017) 381–390.
[37] Y.M. Wang, F.F. Ye, L.H. Yang, Extended belief rule based system with joint learning for environmental governance cost prediction, Ecol. Indic. 111 (2020) 106070.
[38] L.H. Yang, Y.M. Wang, Q. Su, Y.G. Fu, K.S. Chin, Multi-attribute search framework for optimizing extended belief rule-based systems, Inform. Sci. 370-371 (2016) 159–183.
[39] Y.Q. Lin, Y.G. Fu, Q. Su, Y.M. Wang, X.T. Gong, A rule activation method for extended belief rule base with VP-tree and MVP-tree, J. Intell. Fuzzy Systems 33 (6) (2017) 3695–3705.
[40] Y.G. Fu, J.H. Zhuang, Y.P. Chen, L.K. Guo, Y.M. Wang, A framework for optimizing extended belief rule base systems with improved bass trees, Knowl.-Based Syst. 210 (2020) 106484.
[41] L.H. Yang, Y.M. Wang, Y.G. Fu, A consistency analysis-based rule activation method for extended belief-rule-based systems, Inform. Sci. (2018) 445–446, 50-65.
[42] H.Z. Zhu, M.Q. Xiao, L.H. Yang, X.L. Tang, Y.J. Liang, J.F. Li, A minimum centre distance rule activation method for extended belief rule-based classification systems, Appl. Soft Comput. 91 (2020) 106214.
[43] H.Z. Zhu, M.Q. Xiao, X. Zhao, X.L. Tang, L.H. Yang, W.J. Kang, Z.Z. Liu, A structure optimization method for extended belief-rule-based classification system, Knowl.-Based Syst. 203 (2020) 106096.
[44] W.J. Fang, X.T. Gong, G.G. Liu, Y.J. Wu, Y.G. Fu, A balance adjusting approach of extended belief-rule- based system for imbalanced classification problem, IEEE Access 8 (2020) 41201–41212.
[45] L. Dutu, G. Mauris, P. Bolon, A fast and accurate rule-base generation method for Mamdani fuzzy systems, IEEE Trans. Fuzzy Syst. 26 (2) (2018) 715–733.
[46] X.J. Xu, X.P. Yan, C.X. Sheng, C.Q. Yuan, D.L. Xu, J.B. Yang, A belief rule-based expert system for fault diagnosis of marine diesel engines, IEEE Trans. Syst. Man Cybern. Syst. 50 (2) (2020) 656–672.
[47] Z.J. Zhou, G.Y. Hu, C.H. Hu, C.L. Wen, L.L. Chang, A survey of belief rule-base expert system, IEEE Trans. Syst. Man Cybern. Syst. 51 (8) (2021) 4944–4958.
[48] M. Elkano, M. Galar, J. Sanz, H. Bustince, CHI-PG: A fast prototype generation algorithm for big data classification problems, Neurocomputing 287 (2018) 22–33.
[49] M. Elkano, M. Galar, J. Sanz, H. Bustince, CHI-BD: A fuzzy rule-based classification system for big data classification problems, Fuzzy Sets and Systems 348 (2018) 75–101.
[50] Z.J. Zhou, C.H. Hu, J.B. Yang, D.L. Xu, D.H. Zhou, Online updating belief rule based system for pipeline leak detection under expert intervention, Expert Syst. Appl. 36 (4) (2009) 7700–7709.
[51] L.L. Chang, Z.J. Zhou, Y.W. Chen, T.J. Liao, Y. Hu, L.H. Yang, Belief rule base structure and parameter joint optimization under disjunctive assumption for nonlinear complex system modeling, IEEE Trans. Syst. Man Cybern. Syst. 48 (9) (2018) 1542–1554.
[52] Y. Chen, Y.W. Chen, X.B. Xu, C.C. Pan, J.B. Yang, G.K. Yang, A data-driven approximate causal inference model using the evidential reasoning rule, Knowl.-Based Syst. 88 (2015) 264–272.
[53] Y.W. Chen, J.B. Yang, D.L. Xu, Z.J. Zhou, D.W. Tang, Inference analysis and adaptive training for belief rule based systems, Expert Syst. Appl. 38 (10) (2011) 12845–12860.
[54] Z.J. Zhou, C.H. Hu, J.B. Yang, D.L. Xu, M.Y. Chen, D.H. Zhou, A sequential learning algorithm for online constructing belief-rule-based systems, Expert Syst. Appl. 37 (2) (2010) 1790–1799.
[55] Y.M. Wang, L.H. Yang, Y.G. Fu, L.L. Chang, K.S. Chin, Dynamic rule adjustment approach for optimizing belief rule-base expert system, Knowl.-Based Syst. 96 (2016) 40–60.
[56] L.H. Yang, Y.M. Wang, J. Liu, L. Martínez, A joint optimization method on parameter and structure for belief-rule- based systems, Knowl.-Based Syst. 142 (2018) 220–240.

[57] L.L. Chang, Z.J. Zhou, Y.W. Chen, X.B. Xu, J.B. Sun, T.J. Liao, X. Tan, Akaike information criterion-based conjunctive belief rule base learning for complex system modeling, Knowl.-Based Syst. 161 (2018) 47–64.
[58] X.L. Tang, M.Q. Xiao, Y.J. Liang, H.Z. Zhu, J.F. Li, Online updating belief-rule-base using Bayesian estimation, Knowl.-Based Syst. 171 (2019) 93–105.
[59] Y. Guan, Y.-G. Fu, L.-J. Chen, G.-G. Liu, L. Sun, Belief-rule-base inference method based on gradient descent with momentum, IEEE Access 9 (2021) 34487–34499.
[60] W. Zhu, L.L. Chang, J.B. Sun, G.H. Wu, X.B. Xu, X.J. Xu, Parallel multipopulation optimization for belief rule base learning, Inform. Sci. 556 (2021) 436–458.
[61] J.H. Zhuang, J.F. Ye, N.N. Chen, W.J. Fang, X.C. Fan, Y.G. Fu, Extended belief rule-base optimization base on clustering tree and parameter optimization, IEEE Access 9 (2021) 12533–12544.
[62] H. Ishibuchi, T. Yamamoto, T. Nakashima, Hybridization of fuzzy GBML approaches for pattern classification problems, IEEE Trans. Syst. Man Cybern.-Part B: Cybern. 35 (2) (2005) 359–365.
[63] Z. Chen, G. Chen, Building an associative classifier based on fuzzy association rules, Int. J. Comput. Intell. Syst. 1 (3) (2008) 262–273.
[64] J. Alcala-Fdez, R. Alcala, F. Herrera, A fuzzy association rule-based classification model for high-dimensional problems with genetic rule selection and lateral tuning, IEEE Trans. Fuzzy Syst. 19 (5) (2011) 857–872.
[65] G.H. John, P. Langley, Estimating continuous distributions in bayesian classifiers, in: Eleventh Conference on Uncertainty in Artificial Intelligence, San Mateo, 1995, pp. 338-345.
[66] R. Quinlan, C4.5: Programs for Machine Learning, Morgan Kaufmann Publishers, San Mateo, CA.
[67] C.C. Chang, C.J. Lin, LIBSVM: A library for support vector machines, ACM Trans. Intell. Syst. Technol. (TIST) 2 (3) (2011) 1–27.
[68] E. Frank, M.A. Hall, I.H. Witten, The WEKA workbench, in: Online Appendix for Data Mining: Practical Machine Learning Tools and Techniques, Fourth ed., Morgan Kaufmann, 2016.