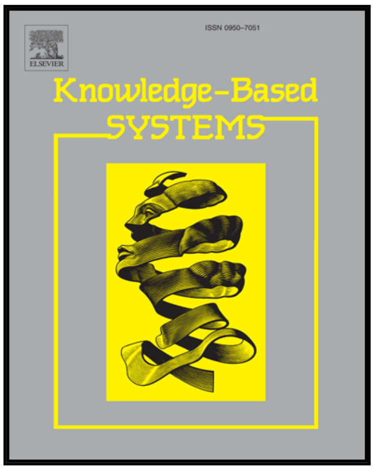
Accepted Manuscript



Akaike Information Criterion-based Conjunctive Belief Rule Base Learning for Complex System Modeling

Leilei Chang , Zhijie Zhou , Yuwang Chen , Xiaobin Xu , Jianbin Sun , Tianjin Liao , Xu Tan

PII: S0950-7051(18)30381-2
DOI: 10.1016/j.knosys.2018.07.029
Reference: KNOSYS 4429

To appear in: *Knowledge-Based Systems*

Received date: 4 February 2018
Revised date: 15 July 2018
Accepted date: 20 July 2018

Please cite this article as: Leilei Chang , Zhijie Zhou , Yuwang Chen , Xiaobin Xu , Jianbin Sun , Tianjin Liao , Xu Tan , Akaike Information Criterion-based Conjunctive Belief Rule Base Learning for Complex System Modeling, *Knowledge-Based Systems* (2018), doi: 10.1016/j.knosys.2018.07.029

Akaike Information Criterion-based Conjunctive Belief Rule Base Learning for Complex System Modeling

Leilei Chang[a, b, c], Zhijie Zhou[c], Yuwang Chen[d], Xiaobin Xu[a], Jianbin Sun[e], Tianjin Liao[f], Xu Tan[b,*]

[a]*School of Automation, Hangzhou Dianzi University, Hangzhou 310018, P. R. China*

[b]*School of Software Engineering, Shenzhen Institute of Information Technology, Shenzhen 518172, P. R. China*

[c] *High-Tech Institute of Xi'an, Xi'an 710025, P R China*

[d]*Decision and Cognitive Science Research Centre, Manchester Business School, The University of Manchester, Manchester M15 6PB, United Kingdom*

[e]*School of Information System and Management, National University of Defense Technology, Changsha 410073, P. R. China*

[f]*State Key Laboratory of Complex System Simulation, Beijing Institute of System Engineering, 10 An Xiang Bei Li Road., Beijing, P R China*

Abstract: Nonlinear complex system modeling has become the basis of many theoretical and practical problems, which requires balancing the correlations between the modeling accuracy and the modeling complexity. However, the two objectives may not be consistent with each other under many practical conditions, especially for complex systems with multiple influential factors. The belief rule base (BRB) has shown advantages in nonlinear complex system modeling under uncertainty. However, most of current works on BRB has focused only on the modeling accuracy. As such, an Akaike Information Criterion (AIC)-based objective, AIC_{BRB}, is deduced to represent both the modeling accuracy (denoted by the Mean Square Error (MSE)) and the modeling complexity (denoted by the number of the parameters). Based on the proposed AIC_{BRB}, a bi-level optimization model and a corresponding bi-level optimization algorithm are developed. Moreover, an empirical optimization path search strategy is proposed for upper-level optimization. The optimization path is comprised of multiple solutions with optimal performance. After the BRB learning process, both the structure and the parameters of BRB are optimized, which identifies the best decision structure of BRB. A numerical multi-extreme function case and a practical pipeline leak detection case are studied. The results show that an optimization path could be identified with a series of optimal solutions. With AIC_{BRB} as the objective, the jointly optimized BRB in the best decision structure can be obtained with an improved modeling accuracy as well as reduced modeling complexity.

Key words: conjunctive belief rule base; Akaike Information Criterion; complex system modeling; optimization path

1. Introduction

With the technological innovations and their impact on operational management, the complexity of many modern systems has far increased [3]. The complex systems in different fields have their unique characteristics, such as nonlinearity, non-differentiability, non-convexity, for which the classical methods fail [18] [26]. Many approaches have been proposed for complex system modeling, such as the Bayesian approaches [29], the agent-based approaches [24], the options-based approaches [3], the uncertainty-based approaches [22] and so on. Among these approaches, the belief rule base (BRB) has been proved to be an efficient tool in addressing the nonlinearity resided in many theoretical and practical complex systems under uncertainty [11] [32] [35] [42].

BRB is comprised of multiple rules in the same belief structure [34] [35]. It has the advantages of generalizing the conventional probability distribution by allowing inexact reasoning [15] and dealing with uncertain information that is different from ignorance and equal likelihood [36]. Moreover, the inference process of BRB is a distributed approximation process which enables BRB to show superior approximation performances [11]. In recent years, BRB has been successfully applied in solving many complex systems related problems [9]-[11] [28]-[42].

However, mostly of present researches on BRB learning for complex systems modeling have employed only the modeling accuracy (denoted by the mean square error (MSE)) as the learning objective [10][11][32]-[36][39]-[43]. Comparatively, other objectives (especially the modeling complexity) have barely been studied, letting alone taking them both under consideration. To jointly optimize of the modeling accuracy and complexity is the primary motive of this study.

Another motive of this study is to identify the optimization path for conjunctive BRB when applied in complex systems modeling. The optimization path in this study denotes a scheme comprised of the referenced values for the attributes with the optimal performance regarding on optimizing systems behavior. The identified optimization path can be used for determining the optimal solution as well as guidance for future modeling and optimization in order to save limited resources and computational time. So far, this topic has not yet been visited.

Driven by the two motives, firstly an integrated criterion is used to represent both the modeling accuracy and the modeling complexity. Among many criterions (such as the Bayesian Information Criterion (BIC) [25] and the Deviance Information Criterion (DIC) [20]), the Akaike Information Criterion (AIC) [1] [2] has been proved to be efficient and effective [4]. Ever since it was proposed, it has been successfully applied in varied fields [1] [5]. In this study, a new AIC-based objective, AIC_{BRB}, is deduced based on the definition of AIC and the characteristics of BRB.

With the new optimization objective AIC_{BRB} being deduced to represent both the modeling accuracy (denoted by the MSE) and the modeling complexity of BRB (denoted by the number of parameters in BRB), rather than simply pursuing the highest modeling accuracy or the lowest modeling complexity.

A bi-level optimization model [12] [26] is constructed based on AIC_{BRB}. For the upper-level model, AIC_{BRB} is used as the optimization objective and the number of the referenced values for the attributes is the decisive parameter. For the lower-level model, the MSE is used as the optimization objective, and the decisive parameters are the referenced values of the attributes, the initial weights of the rules and the beliefs of the scales in the conclusion part of all rules. Correspondingly, a bi-level optimization algorithm is developed. For the upper-level optimization, an empirical optimization path search strategy is applied which is comprised of multiple optimal solutions. For the lower-level optimization, the differential evolutionary (DE) algorithm is employed as the optimization engine [23] [27].

The bi-level optimization model, the corresponding optimization algorithm and the optimization path search strategy constitute the BRB learning approach. After the BRB learning process, both the structure and the parameters of BRB would be optimized and the best decision structure of BRB is derived.

A numerical case and a practical case are studied to verify the efficiency of the BRB learning approach. In the numerical case study, a multiple-extreme non-convex function with five extreme points is modeled. After the BRB learning approach is applied, the optimal result of BRB is identified with six rules instead of the intuitionistic assumed five rules. In the practical case, the modeling accuracy is increased and the modeling complexity is far reduced (the derived BRB is with only 8 rules which is far less than that with 56 rules in previous studies). To conclude, the two cases share the same goal which is to achieve the best decision structure of BRB with the smallest AIC_{BRB}.

Moreover, the empirical optimization path search strategy is also validated. In an extended solution space for the pipeline leak detection case, it is found that to select the (first) local extreme solution as the final optimal solution is practical and more efficient from an engineering perspective, especially in comparison with the global optimization strategy.

The remainder of this study is organized as follows. Literature review is given in Section 2. The basics of BRB and the challenges are introduced in Section 3. AIC_{BRB} is deduced in

Section 4. The optimization model and algorithm are proposed in Sections 5. Two cases are studied in Sections 6 to validate the efficiency of the proposed approach. This study is concluded in Section 7.

2. Literature review

Section 2.1 reviews present studies on BRB parameter learning for complex system modeling. So far, no studies have been conducted on optimization path search. The most related study is the structure learning for BRB which would be reviewed in Section 2.2. Section 2.3 specifically reviews and compares the differences between this study and the related studies which also employed bi-level optimization.

2.1 BRB parameter optimization

So far, most BRB learning approaches focused on parameter optimization which aimed at improving the modeling accuracy (mostly denoted by MSE or MAE).

Yang *et al* proposed the first BRB learning model and algorithm [35]. Later, many researchers proposed related researches with different decisive variables (including initial rule weights, attributes weights, beliefs for scales, referenced values for attributes, etc.) and employs different optimization techniques (including deterministic means and evolutionary algorithms, etc) [10] [11] [21] [32] [42]. Although different models and optimization techniques were adopted, the modeling accuracy was always of main objective, if not the only objective. However, other objectives are also very important, such as the modeling complexity.

Unfortunately, it is difficult to achieve the two goals simultaneously [7] [38]. On the one hand, the modeling accuracy is expected to be high which requires more rules to simulate the nonlinearity caused by the local extreme(s) of the complex system. In a word, high modeling accuracy could invite high complexity. On the other hand, the complexity of the model should be maintained at an appropriate level because the model could not be understood and optimized if it is far too complicated, and it could not accurately represent the nonlinearity and the non-convexity if it is far too simple.

Therefore, the two objectives are inconsistent or even contradicted with each other under many conditions: higher accuracy would likely bring in high complexity, and vice versa. It calls for a new learning approach for conjunctive BRB which jointly optimizes its structure and parameters.

2.2 BRB structure learning

Presently, no studies have been conducted regarding on searching the optimization path. The most related study is BRB structure. Yang *et al* used the Principle Component Analysis

(PCA) to integrate multiple factors into one composite factor (thus a hierarchy is formed) and applied in new product development [36]. Chang *et al* proposed the BRB structure learning approach using multiple dimensionality reduction techniques [8] to select the most representative attributes. Wang *et al* further explored BRB structure learning by deriving techniques based on rough set theory [31]. Li *et al* proposed using conditional generalized minimum variance for feature selection and applied it in safety assessment [19].

However, the conventional BRB structure learning is not the same as the optimization path search as discussed in this study. In conventional BRB structure learning, the goal is either to combine attributes into a composite one or to select more representative attributes. In other words, the ultimate goal for structure learning is always to downsize BRB. While in this study, the goal of optimization path search is to identify a scheme with the optimal performance. With an identified optimization path, a final optimal solution can be obtained.

2.3 Bi-level learning and optimization approaches

Since a bi-level optimization model and algorithm is applied in this study, the related studies are reviewed as follows. For further discussion, readers can refer to Appendix A.

In [26], the evolutionary algorithm for bi-level optimization is proposed based on quadratic approximations (BLEAQ). BLEAQ works by "approximating the optimal solution mapping between the lower level optimal solutions and upper level variables". Comparatively, AIC_{BRB} is used as integrated objective in this study. In other words, it can be analytically calculated by the formula of AIC_{BRB} and requires no evolutionary algorithm such as BLEAQ as in [26].

In [39] [40], the bi-level BRB is constructed denotes that a BRB is constructed with two layers, which is quite different from this study. Moreover, [39] [40] are mostly focused on parameter learning which employed Clonal Section Algorithm (CSA) and Cooperative CoEvolutionary Algorithm (CCEA) respectively and had been achieved superior performances. Comparatively, there was neither structure learning nor joint optimization in [39] [40].

Comparatively, [7] [38] share more in common with this study. They all use one objective for BRB structure and parameter joint optimization. The optimization models are bi-level (even for [38]) in an iterative fashion. However, there are also great dissimilarities between [7] [38] and this study.

In [38], a universal joint optimization approach is proposed for both conjunctive and disjunctive BRBs. Firstly, the optimization objective is different. A generalization error is used in [38] while AIC_{BRB} is used in this study. Furthermore, the structure learning part between the two studied are different. In [38], its structure learning algorithm is based on the heuristic strategy while in this study an optimization path search strategy is proposed. Moreover, [38] can not help identify the optimization path which is a motive of this study. The biggest difference between the two studies is that [38] is universal for both conjunctive and disjunctive BRBs in [38] while this study is specifically designed for conjunctive BRB because the two types of BRBs are assumed different with respective characteristics.

In comparison with [7], this study is primarily designed for conjunctive BRB while [7] is designed for disjunctive BRB. The difference between the two types of BRBs results in the fundamental difference between the two studies. To further explain, conjunctive and disjunctive BRBs are designed for different types of systems and their construction means are different, which thus result in the difference in optimization. To be more specific, for conjunctive BRB, the number of the referenced values for the attributes should be determined first and then a conjunctive BRB can be constructed. Whereas, for disjunctive BRB, its size is firstly determined and then the referenced values for the attributes are determined which constitute a disjunctive BRB. For conjunctive BRB, the attributes are separately optimized and an optimization path can be thus identified. For disjunctive BRB, all of the attributes are simultaneously optimized and there would not be an optimization path.

Although many present studies have been constructed in a "bi-level" fashion, they are different from each other since the characteristics of the search subjects (therefore application conditions and restraints) are different or the model objective/variables are different or the optimization strategies are different. Even for studies such as [7] and this study, they are like two sides of a coin: they share the same material and diameter yet depict two distinctive aspects.

Remark 1: One of the challenges faced by the conjunctive BRB is the "curse of dimensionality" [13], or the combinatorial explosion problem [8] [11]. So far, many endeavors have been made to address this challenge. As introduced above, the structure learning, parameter learning, bi-level or multi-level BRBs can be categorized of three types of endeavors although they are not entirely designed to solve this challenge. Moreover, an online yet incomplete BRB has been made by [41]-[43], and a disjunctive BRB [7] [38] is constructed. As for this study, is it still conducted under the conjunctive BRB. Since the parameters and structure of the conjunctive BRB are jointly optimized, this study also serves as an endeavor to solve the "curse of dimensionality". By applying the disjunctive BRB, this challenge may be further resolved, which would contribute to a new direction for future research.

3. BRB: basics and challenges

3.1 Basics of the belief rule base

BRB is comprised of multiple belief rules in the same belief structure [34]-[35]. The *k*th rule in BRB is described as:

$$\begin{aligned} R_k: \text{if} \quad & \left(A_1 \text{ is } x_1^k\right) \wedge \left(A_2 \text{ is } x_2^k\right) \wedge \cdots \wedge \left(A_M \text{ is } x_M^k\right), \\ & \text{then } \{(D_1, \beta_{1,k}), \cdots, (D_S, \beta_{S,k})\} \\ & \text{with rule weright } \theta_k \end{aligned} \tag{1}$$

where $A_m(m=1,\cdots M)$ denotes the *m*th attribute, $x_m^k(m=1,\cdots M; k=1,\cdots K)$ denotes the referenced values of the *m*th attribute in the *k*th rule, *M* denotes the number of the attributes, $\beta_{s,k}(s=1,\cdots S)$ denotes the belief for the *s*th scale in the *k*th rule, D_s, *S* denotes the number of the scales. $\wedge$' indicates that BRB is constructed under the conjunctive assumption which requires that all combinations of the referenced values of the attributes must be covered.

To summarize, BRB can take in multiple types of information under uncertainty (by the attributes), and transform them (through "IF-THEN") into one unified result in the same belief structure (by the conclusion part). Its efficiency in complex systems modeling has been validated by integrating information from varied sources and in different types [14] [17] [33].

3.2 Challenge and main work of this study

When BRB is applied in complex system modeling, the most important indicator for the performance demonstration is the modeling accuracy which can be represented in multiple forms, such as the (root) mean square error ((R)MSE), the mean abstract error (MAE), the total sum of squares (SST) and so on. The optimization objective of BRB is to maximize the modeling accuracy or to minimize the error, which has been the main objective in present BRB training and learning related researches [9] [11] [31] [32] [35] [42] .

However, another indicator, the modeling complexity represented by the number of the rules, should be taken into consideration as well. The modeling accuracy will grow along with more rules since it can enhance the (local) nonlinearity modeling ability of BRB [9] [31]. However, more rules can also result in high complexity since there will be more parameters and an exponentially growing solution space. Therefore, it requires balancing the correlations between the modeling accuracy and the modeling complexity, which in other words is to achieve the best decision structure of BRB. To do so, three questions need to be addressed.

Question 1: Since the modeling accuracy and the modeling complexity are not consistent with each other under many circumstances, the first question is how to integrate the two objectives into one unified objective.

Question 2: Because there are both discretized (the number of the rules) and continuous (all parameters besides the number of the rules) parameters to be optimized, the second question is how to construct an optimization model guided by the new unified objective.

Question 3: The third question is how to propose a corresponding optimization algorithm to meet the specific requirements of the optimization model.

To sum, ***Question 1*** calls for a unified objective to represent both the modeling accuracy and the modeling complexity. ***Question 2*** calls for an optimization model and ***Question 3*** calls for a corresponding optimization algorithm. The three questions will be discussed in following sections.

4. The Akaike information criterion-based optimization objective

This section intends to propose an integrated objective based on AIC and characteristics of BRB to represent both the modeling accuracy and modeling complexity.

AIC was first proposed by Akaike [1] [2]. For a given linear model, there is

$$z = h_0 + h_1\chi_1 + h_2\chi_2 + \cdots + h_N\chi_N + e \tag{2}$$

where z denotes the output, h_n denotes the input, χ_n denotes the nth model parameter, and e denotes the model noise.

Akaike proposed the following criterion to determine the smallest order (or the number of the independent parameters) for the model.

$$AIC = -2\log L(\hat{\chi}_{ML}) + 2N \tag{3}$$

where $\hat{\chi}_{ML}$ denotes the maximum likelihood estimate (MLE) value of the parameters $\chi = [\chi_1, \chi_2, \cdots, \chi_N]$, $L(\hat{\varphi}_{ML})$ denotes the likelihood function under $\hat{\chi}_{ML}$, and N denotes the number of the independent parameters in the model. There are two parts in (3). The first part of the model, $-2\log L(\hat{\chi}_{ML})$, favors the model with a higher modeling accuracy while the second part, $2N$, favors the model with fewer parameters. The model with the smallest AIC should be selected as the optimal model.

Originally, AIC was designed and proposed for the linear model. However, the input(s) and the output of BRB are in the form of the “IF-THEN” rule instead of linear analytical relationships [34] **Error! Reference source not found.**. Moreover, the inference mechanism of BRB is a distributed approximation process which is not necessarily derivable and monotonic [11]. Therefore, BRB is nonlinear by definition and with unique characteristics, which makes AIC could not be directly applied for BRB learning.

Next, the AIC-based objective AIC_{BRB} will be deduced [6]. Assume that $\{(\mathbf{X},\mathbf{Y})\}$ is the training dataset where $\mathbf{X}$ is the input and $\mathbf{Y}$ is the output. $\mathbf{X}$ has two dimensions. One is the number of the training dataset, P, and the other is the number of the independent parameters, N. So there is $\mathbf{X}=[\mathbf{X}_n^p, n=1,\cdots,N; p=1,\cdots,P]$. $\mathbf{Y}$ has one dimension which is the number of the training dataset. So there is $\mathbf{Y}=[y_1, y_2,\cdots,y_P]^T$. The output of BRB can be modeled by

$$f(\mathbf{X}^p)=\omega_0+\sum_{n=1}^{N}\omega_n\phi_n(\mathbf{X}_n^p) \tag{4}$$

where ω_n denotes the weight of the *n*th independent parameter, $n=1,2,\cdots,N$, and $\phi_n(\mathbf{X}_n^p)$ denotes the mapping correlations between the input and the estimated output of BRB regarding on the *n*th independent parameter and the *p*th set of the training data. $f(\mathbf{X}^p)$ denotes the estimated output of BRB with the input $\mathbf{X}^p$.

Let ε_p be the error between $f(\mathbf{X}^p)$ and y_p. Assume that ε_p follows the normal distribution, $\varepsilon_p \sim N\{0,\sigma^2\}$. Eq. (4) can be written as

$$y_p=f(\mathbf{X}^p)+\varepsilon_p=\omega_0+\sum_{n=1}^{N}\omega_n\phi_n(\mathbf{X}_n^p)+\varepsilon_p \tag{5}$$

Based on Eq. (5), there is $y_p \sim N(\omega_0+\sum_{n=1}^{N}\omega_n\phi_n(\mathbf{X}_n^p),\sigma^2)$. The likelihood function of Eq. (5) can be further written as

$$\begin{aligned} L(\mathbf{Y},\mathbf{W},\sigma^2)&=\prod_{p=1}^{P}\frac{1}{\sqrt{2\pi}\sigma}\exp\{-\frac{1}{2\sigma^2}(y_p-\omega_0-\sum_{n=1}^{N}\omega_n\phi_n(\mathbf{X}_n^p))^2\} \\ &=(2\pi\sigma^2)^{-\frac{P}{2}}\exp\{-\frac{1}{2\sigma^2}(y_p-\omega_0-\sum_{n=1}^{N}\omega_n\phi_n(\mathbf{X}_n^p))^2\} \end{aligned} \tag{6}$$

where $\mathbf{W}=[\omega_1,\omega_2,\cdots,\omega_N]^T$.

By using the log transformation, Eq. (6) can be written as

$$\ln L=-\frac{P}{2}\ln(2\pi)-\frac{P}{2}\ln\sigma^2-\frac{1}{2\sigma^2}\sum_{p=1}^{P}(y-\omega_0-\sum_{n=1}^{N}\omega_n\phi_n(\mathbf{X}_n^p))^2 \tag{7}$$

By deriving the partial derivatives of ω_n and σ^2 on $\ln L$ based on Eq. (7), let the partial derives be equal to 0, and the canonical equations are in Eq. (8)

$$\begin{cases} \dfrac{\partial \ln L}{\partial \omega_n}=0 \\ \dfrac{\partial \ln L}{\partial \sigma^2}=0 \end{cases} \tag{8}$$

where $n = 1, 2, \cdots, N$.

By solving Eq. (8), the solution is as follows

$$\mathbf{W} = [\omega_1, \omega_2, \cdots, \omega_N]^T$$

The maximum likelihood estimation for $\mathbf{W}$ and σ^2 can be calculated by,

$$\mathbf{W} = (\mathbf{G}'\mathbf{G})^{-1}\mathbf{G}'\mathbf{Y} \tag{9}$$

$$\sigma^2 = \frac{1}{P}(\mathbf{Y} - \mathbf{GW})'(\mathbf{Y} - \mathbf{GW}) \tag{10}$$

where $\mathbf{G} = \begin{bmatrix} 1 & \phi_1(X_1^1) & \cdots & \phi_N(X_N^1) \\ 1 & \phi_1(X_1^2) & \cdots & \phi_N(X_N^2) \\ \vdots & \vdots & & \vdots \\ 1 & \phi_1(X_1^P) & \cdots & \phi_N(X_N^P) \end{bmatrix}$

Since Eqs. (9-10) are in the matrix forms. Eq. (7) could be written as

$$\ln L(\mathbf{Y}, \mathbf{W}, \sigma^2) = -\frac{P}{2}\ln(2\pi) - \frac{P}{2}\ln\sigma^2 - \frac{P}{2} \tag{11}$$

With Eq. (11), Eq. (3) could be presented as

$$AIC_{BRB} = -2[-\frac{P}{2}\ln(2\pi) - \frac{P}{2}\ln\sigma^2 - \frac{P}{2}] + 2N \tag{12}$$

Eq. (12) can be further written as

$$AIC_{BRB} = P\ln\sigma^2 + 2N + C \tag{13}$$

where $C = P\ln(2\pi) + P$ is a constant and is unrelated to N.

The constant, C, could be omitted in comparison among different models. Therefore, Eq. (13) could be written as

$$AIC_{BRB} = P\ln\sigma^2 + 2N \tag{14}$$

In Eq. (14), σ^2 can be calculated by

$$\sigma^2 = P \bullet MSE \tag{15}$$

Based on Eq. (14) and Eq. (15), Eq. (13) is transformed as

$$AIC_{BRB} = P\ln(P \bullet MSE) + 2N \tag{16}$$

From Eq. (16), the first part of AIC_{BRB} is the MSE which denotes the modeling accuracy, and the second is the number of the parameters which denotes the modeling complexity.

Note that AIC_{BRB} and the MSE are related. When there are fewer parameters, AIC_{BRB} is sensitive to the improvement of the MSE, as explained by Fig. 1.

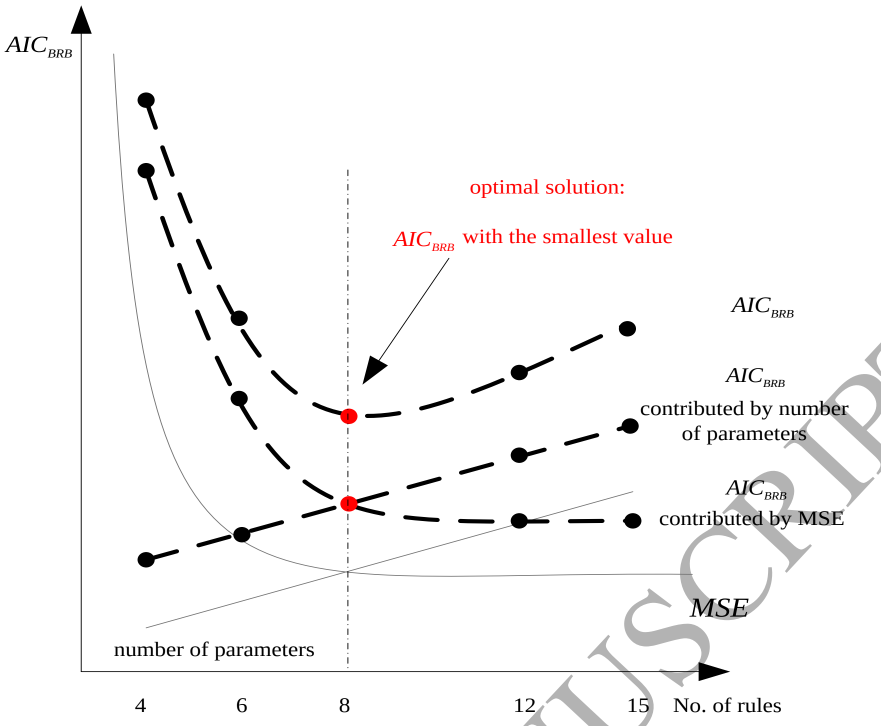


Fig. 1 Illustration of identifying the optimal solution based on AIC_{BRB}

In Fig. 1, in the left of the dashed line or when there are less than 8 rules, AIC_{BRB} and the MSE both decrease while the number of the rules/parameters increases because the influence of the MSE outweighs that of the number of the parameters. When BRB is beyond a certain size, as depicted in the right of the dashed line or when there are more than 8 rules in Fig. 1, although the MSE still slowly decreases, AIC_{BRB} starts to increase because the influence of the number of the parameters outweighs that of the MSE. Therefore, the BRB with the smallest AIC_{BRB} should be identified as the optimal solution, as is the red point or 8 rules in Fig. 1.

Remark 2: AIC_{BRB} in Eq. (16) and Fig. 1 is consistent with the original definition of AIC in Eq. (3). Moreover, it is also consistent with the relationship between the modeling accuracy and the modeling complexity. When there are fewer parameters, the model is not complex yet it would also be less accurate. However, more parameters would likely lead to higher accuracy as well as more complexity. It thus calls balancing between the modeling accuracy and complexity via optimizing AIC_{BRB}.

Remark 3: [28] has proposed a similar idea by using AIC in BRB structure validation. The main differences of this study in comparison with [28] are: (1) detailed deduction process of AIC_{BRB} is given (in this section) based on the characteristics of BRB, (2) a BRB learning approach comprised of an optimization model with AIC_{BRB} as the optimization objective and the corresponding optimization algorithm is proposed in Sections 4 and 5, (3) an optimization path search strategy is proposed in Section 5.3, and (4) both the structure and parameters are optimized and the best decision structure of BRB is derived. These differences are also the main contributions of this study.

5. The Akaike information criterion-based optimization model and algorithm

5.1 BRB learning approach framework

With new objective, a new BRB learning approach is proposed with a bi-level optimization model and a corresponding optimization algorithm.

For the upper-level optimization model and algorithm, it is designed to implement the optimization search strategy via optimizing the number of referenced values for the attributes. For the lower-level optimization model and algorithm, it is designed to optimize the parameters with given number of referenced values for the attributes. The local optimization is designed to preliminarily check if it is worthy of continuing the optimization process. The upper- and lower-level optimization is implemented in an iterative fashion to identify the optimal configuration of the structure and parameters for a conjunctive BRB.

The framework of the new BRB learning approach is illustrated in Fig. 2.

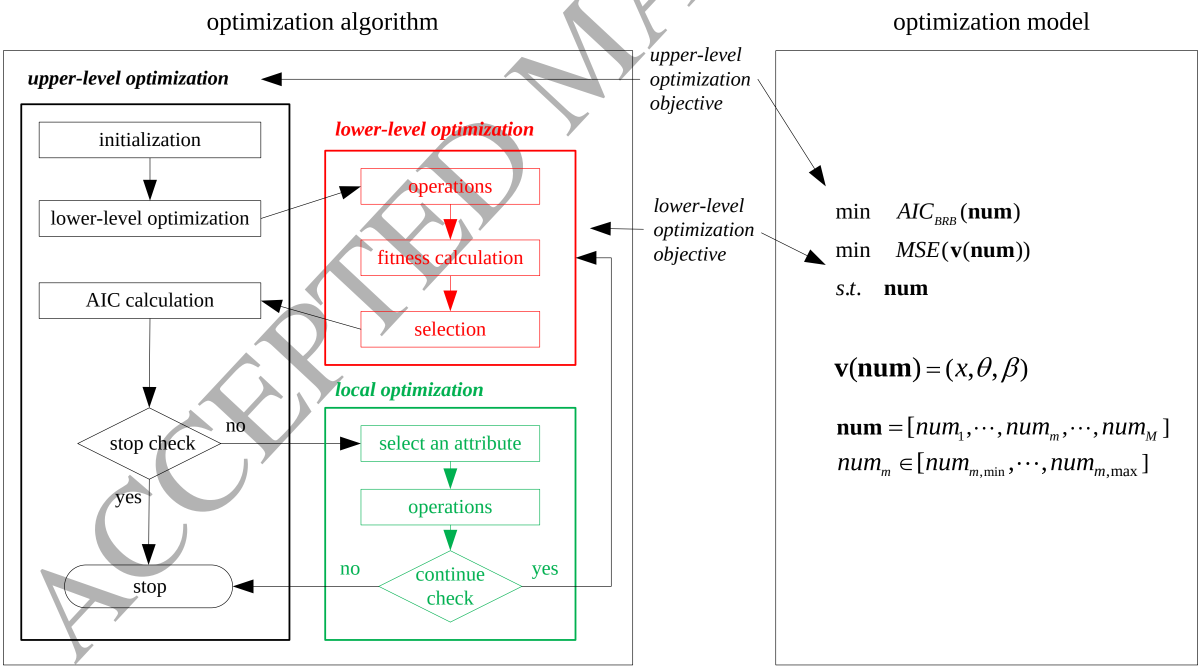


Fig. 2 new BRB learning approach framework

5.2 Bi-level optimization model

In this section, the optimization model is constructed with AIC_{BRB} as the optimization objective. As in Eq. (16), AIC_{BRB} is comprised of two parts: the MSE and the number of the parameters. The concept of the MSE is clear while the number of the parameters needs to be further analyzed. When the number of the attributes and the number of the scales in the conclusion part are fixed, the number of the parameters is determined only by the number of the referential values for the attributes, $\mathbf{num} = (num_m \mid m = 1, \cdots, M)$. Moreover, besides that the number of the referential values for the attributes is discrete, the other parameters are all continuous, namely the referenced values of the attributes, the initial weights of the rules and the beliefs of the scales in the conclusion part. Therefore, the bi-level optimization model [12] [26] is constructed as follows.

$$\begin{aligned} &\min \quad AIC_{BRB}(\mathbf{num}) \\ &\min \quad MSE(\mathbf{v}(\mathbf{num})) \\ &s.t. \quad \mathbf{num} \end{aligned} \tag{17}$$

For the lower-level optimization model, the optimization objective is the MSE and the decisive parameters form a vector, $\mathbf{v}(\mathbf{num})$, which are the referenced values of the attributes, the initial weights of the rules and the beliefs of the scales in the conclusion part. For the upper-level optimization model, the optimization objective is AIC_{BRB}, and decisive parameter is the number of the referenced values for the attributes, $\mathbf{num} = (num_1, num_2, \cdots, num_M)$.

5.2.1 Lower-level optimization model with the MSE as optimization objective

For the lower-level optimization model, the objective is the MSE between the actual outputs and the estimated outputs of the model. The decisive parameters are as follows.

(1) The referenced values of *M* attributes, x. For the *m*th attribute in the *k*th rule, the referenced values should be within the range of the lower and upper bounds, $lb_m \le x_m^k \le ub_m$. Moreover, the first and the last referenced values of the *m*th attribute should be equal to the lower and the upper bounds, $x_m^i = lb_m$, $x_m^K = ub_m$. Since there are at least two referenced values (the lower and the upper bounds) for the *m*th attribute, $num_m \ge 2$. The number of x to be optimized is $\sum_{m=1}^{M}(num_m - 2)$.

(2) The initial weights of the rule, θ. The initial weight of the *k*th rule θ_k should be within the range of (0, 1], $0 < \theta_k \le 1$. Note that θ_k is not the final weight of the *k*th rule, w_k, which still needs to be multiplied by the activated weight $w_{activated,k}$, $w_k = \theta_k * w_{activated,k}$. The number of θ to be optimized is K.

(3) The belief of each scale in each rule, β. The belief of the *s*th scale in the *k*th rule should be within the range of [0, 1], $0 \le \beta_{s,k} \le 1$. The sum of all the beliefs in the *k*th rule

should less or equal to 1, $\sum_{s=1}^{S}\beta_{s,k} \le 1$ ($\sum_{s=1}^{S}\beta_{s,k} < 1$ when there is incomplete information). The number of β to be optimized is $K*S$.

Since BRB is constructed under the conjunctive assumption, the number of the rules in BRB is $\prod_{m=1}^{M} num_m$. Then the total number of the parameters is

$$Num_{para} = \sum_{m=1}^{M}(num_m - 2) + K + K*S \tag{18}$$

Based on Eq. (18), it can be observed that Num_{para} is mostly determined by the number of rules K and the number of scales S.

When the referenced values of the attributes are fixed, the number of the rules in BRB is determined. For a BRB with K rules, the decisive parameters form a vector, $\mathbf{v}(\mathbf{num}) = (x\ ,\theta\ ,\beta\ )$.

The lower-level optimization model is given as follows.

$$\min MSE(x,\theta,\beta) \tag{19}$$

s.t.

$$lb_m \le x_m^k \le ub_m \tag{19a}$$

$$x_m^1 = lb_m \tag{19b}$$

$$x_m^K = ub_m \tag{19c}$$

$$0 < \theta_k \le 1 \tag{19d}$$

$$0 \le \beta_{s,k} \le 1 \tag{19e}$$

$$\sum_{s=1}^{S}\beta_{s,k} \le 1 \tag{19f}$$

where (19a/b/c) denote that the referenced values for the attributes should be within the range of the lower and upper bounds of the attributes; (19e) denotes that the initial weights for the attributes should be within (0,1]; (19e) denotes that the beliefs of the scales in the conclusion part should be within [0, 1]; (19f) denotes that the sum of beliefs of the scales in the conclusion part should be equal or smaller than 1.

5.2.2 Upper-level optimization model for optimization path search

The upper level optimization is designed specifically for conjunctive BRB to search the optimization path. For the upper-level optimization model, AIC_{BRB} is the optimization objective and the decisive parameter is the number of the referenced values of the attributes which determines the number of the rules/parameters in BRB.

The upper-level optimization model is as follows.

$$\min \; AIC_{BRB}(\mathbf{num}) \tag{20}$$

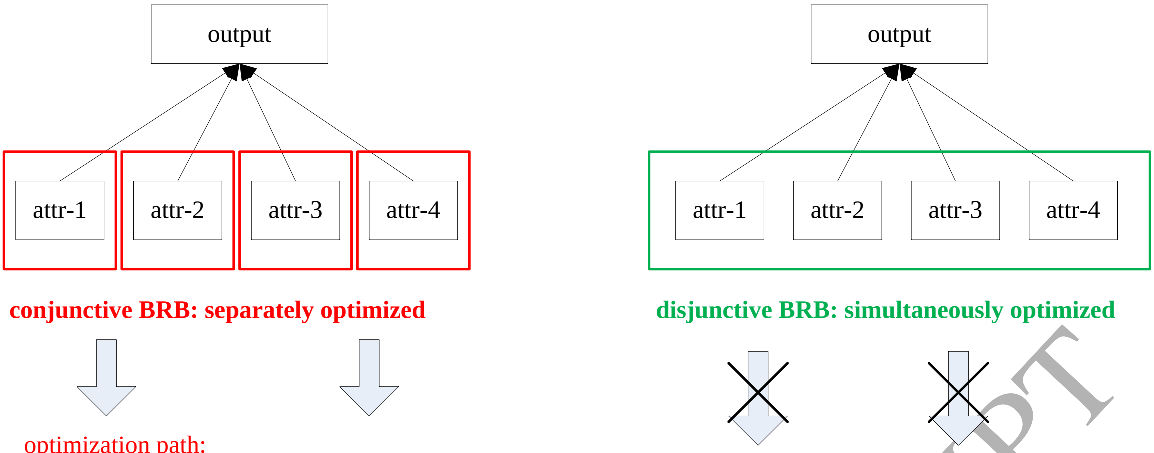
s.t.

$$num_{m,\min} \le num_m \le num_{m,\max};\, m \in [1,\cdots,M] \tag{20a}$$

where $num_{m,min}$ and $num_{m,max}$ are the minimum/maximum number of the referenced values of the *m*th attribute, respectively. Normally, $num_{m,min}$ for any attribute is set at 2 since an attribute needs at least two referenced values for hold the lower and upper bounds. As for $num_{m,max}$, it is normally set at a big number (i. e., 20 or 30) but the optimization process would likely stop at a local optimal solution before it reaches $num_{m,max}$.

After upper-level optimization, the BRB learning process is terminated as well. An optimization path would be completed as it is the integration of the optimal solutions. With AIC_{BRB} as the integrated objective, a final optimal solution can be determined. Moreover, an optimization path can also be used as guidance for future modeling and optimization in order to save computational resources and time.

Remark 4: The upper level model for optimization path search is only applicable for conjunctive BRB (inapplicable for disjunctive BRB). This is because the referenced values for the attributes are separately optimized for conjunctive BRB while they are simultaneously optimized for disjunctive BRB, as shown in Fig. 3. To further explain, a conjunctive BRB is constructed only if the referenced values for the attributes are firstly determined. For a disjunctive BRB, a new disjunctive rule can be inserted into a present disjunctive BRB as long as it can improve modeling accuracy. This is fundamental difference between the conjunctive and disjunctive BRB which also result in that [7] and this study are different from each other.



$$(x_1^*, x_2^*, x_3^*, x_4^*) = [(x_1^1, x_2^1, x_3^1, x_4^1), (x_1^1, x_2^2, x_3^1, x_4^1), (x_1^2, x_2^2, x_3^1, x_4^1), \ldots\ldots (x_1^{k_1}, x_2^{k_2}, x_3^{k_3}, x_4^{k_4})]$$

Fig. 3 Upper-level optimization model for optimization path search ($(x_1^{k_1}, x_2^{k_2}, x_3^{k_3}, x_4^{k_4})$ denotes different referenced values for attributes in different optimal solutions)

5.3. Bi-level optimization algorithm guided by the bi-level optimization model

5.3.1 Lower-level optimization algorithm for parameter learning

For the lower-level optimization model in Section 5.2.1, the corresponding lower-level optimization algorithm is proposed by using the differential evolutionary algorithm (DE) [27] as the optimization engine. The main steps are as follows:

Step 1: Parameters initialization.

With the number of the referenced values of the attributes (determined in the upper-level algorithm), **num**, the initial parameters are the parameters for BRB and DE.

The parameters for BRB are x, θ, and β, which must meet the restraints given in Eq. (19a-f). The parameters for DE are the size of the population, *size_pop*, the number of the generation, *num_gen*, and the upper and lower boundaries of the attributes/weights/scales. Details of the coding of the parameters are in Appendix B.

Step 2: Optimization operations.

The conventional DE is applied as the optimization engine. Details of the crossover and the mutation operations are in Appendix B.

Step 3: Fitness function.

Step 3.1: The rule activation, matching degree calculation and weight calculation procedures are given in Appendix C.

Step 3.2: ER inference.

After certain rules being activated, the activated L rules are integrated using ER, and the analytic form of ER is given in Eqs. (21)-(22) [30],

$$\beta_s = \frac{\mu[\prod_{k=1}^{L}(w_k\beta_{s,k} + 1 - w_k\sum_{s=1}^{S}\beta_{s,k}) - \prod_{k=1}^{L}(1 - w_k\sum_{s=1}^{S}\beta_{s,k})]}{1 - \mu[\prod_{k=1}^{L}(1 - w_k)]} \tag{21}$$

$$\mu = [\sum_{n=1}^{N}\prod_{k=1}^{L}(w_k\beta_{n,k} + 1 - w_k\sum_{s=1}^{N}\beta_{s,k}) - (N-1)\prod_{k=1}^{L}(1 - w_k\sum_{s=1}^{N}\beta_{s,k})]^{-1} \tag{22}$$

where β_s represents the belief for the sth scale.

Step 3.3: Output by utility.

The utility of the sth scale D_s is denoted as $U\left(D_s\right)$. The integrated utility of the belief distribution is given in Eq. (23).

$$T = \sum_{s=1}^{S} U\left(D_s\right)\beta_s \tag{23}$$

Step 4: Selection

The selection operation is also in Appendix B.

Step 5: Stop criterion check.

Check on whether the stop criterion (which is normally the number of generation, e. g. 500 generations) has been met. If not, go to Step 3; otherwise, the individual with the smallest MSE will be selected as the optimal solution.

5.3.2 Upper-level optimization algorithm for optimization path search

For the upper-level optimization model given in Section 5.2.2, the corresponding upper-level optimization algorithm for optimization path search is illustrated in Fig. 4. The optimization path search strategy is an iterative process. In each iteration, the performance of BRB with a newly added referenced value for an attribute is compared with that of the previous BRB.

Unlike conventional structure learning which intends to downsize BRB or parameter learning which aims to improving modeling accuracy, an optimization path is a scheme comprised of selected referenced values for the attributes with the optimal performance regarding on modeling complex system's behavior.

In Fig. 4, with R_{11} as the supposed starting rule, the optimization strategy searches each attribute. When a new rule is found to be with a better performance, that rule would be identified as the next optimization rule. By extending this strategy, an optimization path would be identified as marked red in Fig. 4.

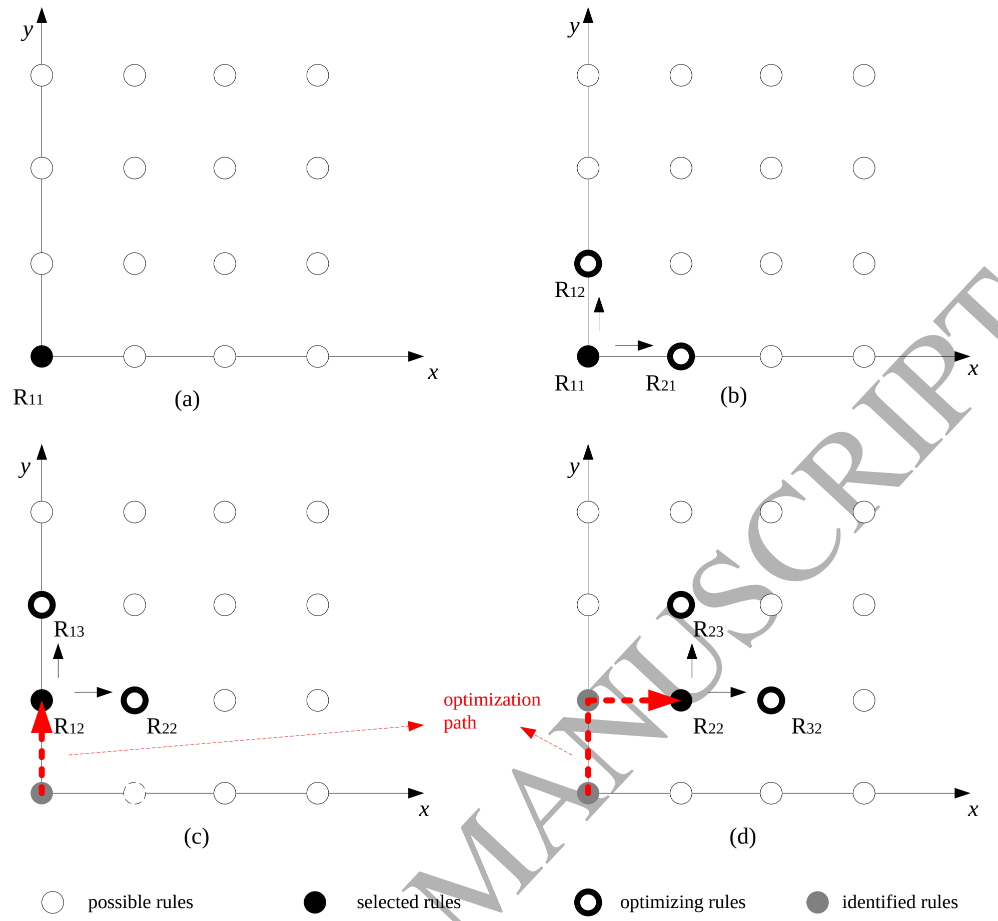


Fig. 4 Optimization path search upper-level optimization

The detailed steps of the upper level optimization algorithm are given as follows:

Step 1: Initialization. For the upper-level optimization, the parameters are the initial number of the referenced values for each attribute.

Step 2: Go to the lower-level optimization with the initialized parameters.

Step 3: Calculate AIC_{BRB} by Eq. (16) with the number of the parameters and the MSE derived from the lower-level optimization algorithm.

Step 4: Check AIC_{BRB}. For the mth attribute,

if $num_m > num_{m,ini}$

  Compare $AIC_{BRB}(m, num_m - 1)$ with $AIC_{BRB}(m, num_m)$;

  if $AIC_{BRB}(m, num_m - 1) < AIC_{BRB}(m, num_m)$

    stop;

  end

else

  go to *Step* 6;

end

Step 5: Stop criterion check. For the mth attribute,

if $num_m = num_{m,\max}$
    stop;
else
    go to *Step* 6;
end

Step 6: Optimization path search as given in Section 5.4.

Remark 5: Since the first order derivative of AIC_{BRB} is non-monotonic, there are multiple local extreme points. However, within a small interval, the first order derivative of AIC_{BRB} could be recognized as monotonic. Therefore, an empirical strategy is applied by comparing $AICs_{BRB}$ for each attribute to determine if the (first) local extreme point is derived.

Remark 6: The computational complexity of the upper level optimization algorithm is rather low because it is assumed in this study that the first local optimal extreme would be selected as the final solution. For a conjunctive BRB with *M* attributes, the number of referenced values needs to be compared is also *M*. Even for conservative reassurance by comparing two referenced values for each attribute, the total computation would be 2*M*. To summarize, the computation with n_m referenced values for the *m*th attribute calculated would be $\sum_{m=1}^{M} n_m$.

5.3.3 Local optimization

After lower-level optimization, a local optimization strategy is introduced to determine whether or not it should go into the next iteration of bi-level optimization. In the local optimization, a new referenced value is introduced to each attribute. Thus the local optimization is targeted with only newly added rules (generated by newly added referenced values) and its neighbor rules. Moreover, it is also assumed with shorter generations.

Detailed steps of the local optimization are given as follows:

Step 1: Select the *m*th attribute and keep the other attributes fixed.

Step 2: For the *m*th selected attribute, randomly select two neighbor referenced values.

Step 3: Local optimization.

Step 2.1: Parameters initialization. See Step 1 in Section 5.1.

Step 2.2: Crossover and mutation. See Step 2 in Section 5.1.

Step 2.3: Fitness calculation. See Step 3 in Section 5.1.

Step 2.3.1: Rule activation, matching degrees and weights calculation.

Step 2.3.2: ER inference.

Step 2.3.3: Output by utility.

Step 2.4: Selection. See Step 4 in Section 5.1.

Step 2.5: Stop criterion check. See Step 5 in Section 5.1. Note that the stop criterion is usually less strict. Normally, 50 generations would be an appropriate choice.

Step 4: Calculate $AIC_{BRB}(\mathbf{num})$. See Eq. (16) in Section 3.

Step 5: Select the next attribute and repeat Steps 2 to 4 until all attributes are locally optimized.

Step 6: Compare different $AIC_{BRB}(\mathbf{num})$ and determine the optimization path. Assume that there are two attributes, *m* and *q*,

Compare $AIC_{BRB}((m, num_m), (q, num_q))$
    with $AIC_{BRB}((m, num_m + 1), (q, num_q)), AIC_{BRB}((m, num_m), (q, num_q + 1))$
if $AIC_{BRB}((m, num_m + 1), (q, num_q))$ is the smallest
    attribute m is the optimization direction
else
    if $AIC_{BRB}((m, num_m), (q, num_q + 1))$ is the smallest
        attribute q is the optimization direction
    else
        if $AIC_{BRB}((m, num_m), (q, num_q))$ is the smallest
            $(m, num_m), (q, num_q)$ is the optimial solution
        end
    end
end

Step 7: Stop criterion check. If a smaller AIC_{BRB} is found, stop; otherwise, go to Step 3.

6. Case studies

6.1 A numerical case study: multi-extremal non-convex function

In this numerical case study, the following multi-extremal non-convex function which has been studied in [11] is as follows.

$$g(x) = e^{-(x-2)^2} + 0.5e^{-(x+2)^2}, -5 \le x \le 5 \tag{26}$$

The settings of this case are: 1) BRB is initiated from 3 rules and stopped at 50 rules (or a large number of the rules would suffice), 2) there are 5 scales in the conclusion part with the utilities as $\{D_1, D_2, D_3, D_4, D_5\} = \{-0.5, 0, 0.5, 1, 1.5\}$which is the same as in [11], 3) 20 individuals, 4) 2000 generations for lower-level optimization and 50 generations for optimization path search, 5) this experiment is performed with 30 runs, 6) the training/testing dataset is the same as that of [11], 101 sets of data which are uniformed sampled from the interval [-5, 5], 7) and the initial BRB is randomly generated because DE is applied as the optimization engine.

Based on the observation of the function of (26) in Fig. 5, there are five distinct extremes points which are the two ends and three local extreme points (around "-2", "0" and "2" in the x-axis direction). Thus, it is intuitionistic to assume BRB should be of 5 rules which is the case in [11].

However, the statistical results in Table 1 unveil a different result, which indicated that BRB with six rules has been identified as the optimal solution in 23 runs out of 30. The AIC_{BRB} statistics of the derived BRBs in the best decision structure of all 30 runs are shown in Table 2. BRB with five and six rules are shown in Table 3 and Table 4 with MSEs of 4.00E-05 and 1.70E-05, respectively. Comparatively, they are 36.71% and 73.10% less than 6.32E-05 of [11]. The estimated results and the error using the BRB in [11] and the BRBs with five/six rules are shown in Fig. 5.

Table 1 No. runs with different No. rules identified as the optimal solution for Case I

No. rules identified as the optimal solution	4	5	6	7	8
No. runs	1	4	23	0	2

Table 2 statistics of AIC_{BRB} for Case I optimized by DE

	min	average	variance
AIC_{BRB}	-199.19	-171.52	120.19

Table 3 BRB with five rules for Case I

No.	Rule weight	x	Belief degrees on $\{D_1, D_2, D_3, D_4, D_5\}$
1	0.6695	-5.0000	(0.4389 0.3458 0.0016 0.2129 0.0008)
2	0.2347	-2.1615	(0.0020 0.6714 0.0011 0.0002 0.3252)
3	0.2787	-0.26012	(0.7262 0.0101 0.0017 0.0001 0.2619)
4	0.1655	2.0347	(0.0005 0.0046 0.0137 0.9800 0.0012)
5	0.4488	5.0000	(0.1974 0.6242 0.1771 0.0001 0.0012)

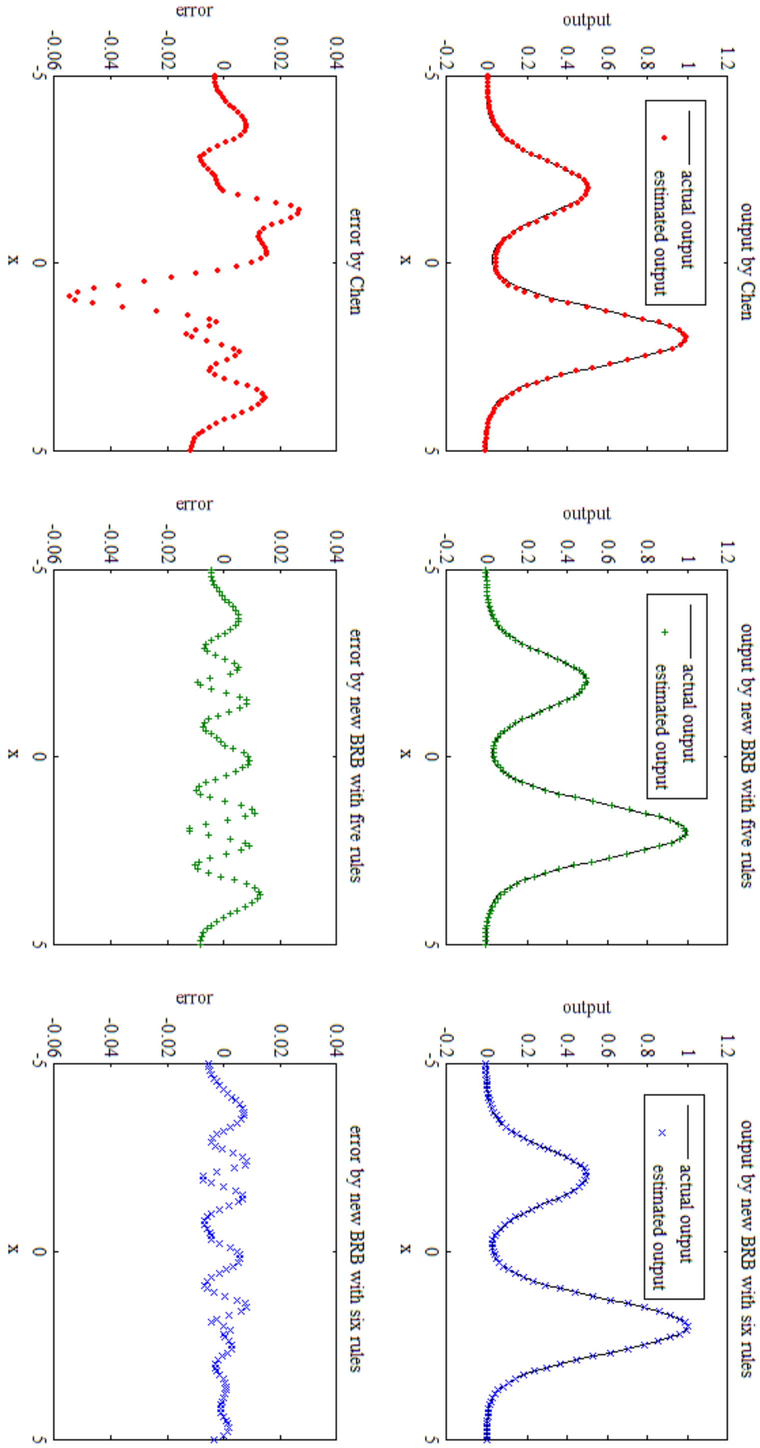


Fig. 5 Comparison on outputs and errors of the estimated and actual results for Case I

Table 4 BRB with six rules for Case I

No.	Rule weight	x	Belief degrees on {D_1, D_2, D_3, D_4, D_5}
1	0.6377	-5.0000	(0.3941 0.3095 0.2391 0.0282 0.0292)
2	0.2216	-2.1289	(0.0004 0.5045 0.0063 0.4888 0.0001)
3	0.2687	-0.25879	(0.5792 0.1048 0.0025 0.3128 0.0007)
4	0.1635	2.0536	(0.0001 0.3161 0.0017 0.0427 0.6394)
5	0.3010	4.4346	(0.1504 0.7696 0.0050 0.0738 0.0012)
6	0.0076	5.0000	(0.6462 0.0194 0.0303 0.3038 0.0003)

By analyzing the referenced values for the attributes, it is discovered that the referenced values of the BRB with five rules derived in Table 3 is quite similar with that by [11]. However, the MSE has decreased by 36.71% from 6.32E-05 to 4.00E-05.

The BRB with six rules in Table 4 indicates that one new referenced value, 4.4346, has been added. Although this new referenced value may not be as representative as the other five referenced values in Table 3 and Fig. 3, it can improve the nonlinearity modeling ability because the MSE of BRB in Table 4 has been decreased from 4.00E-05 to 1.70E-05 by 57.50%. Moreover, AIC_{BRB} is -199.19 which is smaller than -175.76 of the BRB with five rules in Table 3 by 13.33%.

Remark 7: The best decision structure of BRB for Case I is finally identified with six rules instead of the intuitionally assumed five rules (the upper and lower boundaries and three local extreme points in Fig. 3). This proves that, by using AIC_{BRB} as the optimization objective, the BRB learning approach only stops when the balance between the modeling accuracy and the modeling complexity is achieved, rather than simply pursuing a higher modeling accuracy or a lower modeling complexity.

6.2 A practical case study: pipeline leak detection

6.2.1 Background and parameter settings

A part of a pipeline in northern UK has shown leakage and data are collected during a leakage trial [32]. Two factors, the pressure difference and the flow difference, namely *FlowDiff* and *PressureDiff*, have been selected as two factors (attributes) to construct a BRB model for the pipeline leak detection case. So far, this case has been studied in multiple papers so that it can be used as a benchmark case.

The parameter settings of this case are as follows:

1) BRB is initiated from 4 rules with 2 referenced values for each attribute,

2) There are 5 scales in the conclusion part with the utilities as {D_1, D_2, D_3, D_4, D_5} = {0, 2, 4, 6, 8}which is the same as in previous studies,

3) 20 individuals,

4) 500 generations for lower-level optimization and 50 generations for optimization path search,

5) This experiment is performed with 30 runs,

6) 500 sets of data are randomly selected as the training dataset and the total dataset is used as the testing dataset,

7) The initial BRB is randomly generated because DE is applied as the optimization engine.

6.2.2 Case study result

Table 5 shows the derived BRB with eight rules as the optimal solution and the best decision structure. In comparison with most of previous studies which intuitionally selected eight referenced values for *FlowDiff* and seven referenced values for *PressureDiff* (BRB is of 56 rules), the BRB in Table 5 can produce an improved result with four referenced values for *FlowDiff* and two referenced values for *PressureDiff* (BRB is of 8 rules), respectively. The estimated and actual outputs, along with the error between them at each time point by this study and [42], are shown in Fig. 6.

Table 5 BRB with eight rules for Case II

	weight	*FlowDiff*	*PressureDiff*	Conclusion
1	0.9799	-10.0000	-0.0200	0.9944, 0.0000, 0.0023, 0.0032, 0.0000
2	0.9833	-10.0000	0.0400	0.9818, 0.0033, 0.0026, 0.0123, 0.0000
3	0.4990	-7.1513	-0.0200	0.1921, 0.0989, 0.2383, 0.0027, 0.4680
4	0.6651	-7.1513	0.0400	0.0575, 0.0237, 0.0254, 0.0064, 0.8870
5	0.0309	-1.3953	-0.0200	0.6932, 0.0001, 0.2862, 0.0027, 0.0179
6	0.0228	-1.3953	0.0400	0.7734, 0.2204, 0.0028, 0.0026, 0.0008
7	0.5780	2.0000	-0.0200	0.9708, 0.0254, 0.0001, 0.0022, 0.0016
8	0.0297	2.0000	0.0400	0.5692, 0.1947, 0.1065, 0.0796, 0.0500

Note: The sum of the beliefs in the conclusion part in each rule may not be equal to "1" since all numerical values are rounded to 4 decimal places in MATLAB programming, e. g., the sum of the beliefs in rule 1 is 0.9999, and the sum of the beliefs in rules 5 and 7 is 1.0001.

Table 6 shows the comparison with present researches. It shows that, this study produces the smallest MSE and AIC_{BRB} for conjunctive BRB in comparison with previous studies mainly because rules are far reduced (the MSEs are reduced as well). In retrospect, maybe the conjunctive BRBs in previous studies suffered from overfitting. Only [9] which constructed a disjunctive BRB shows superior performance in comparison with this study.

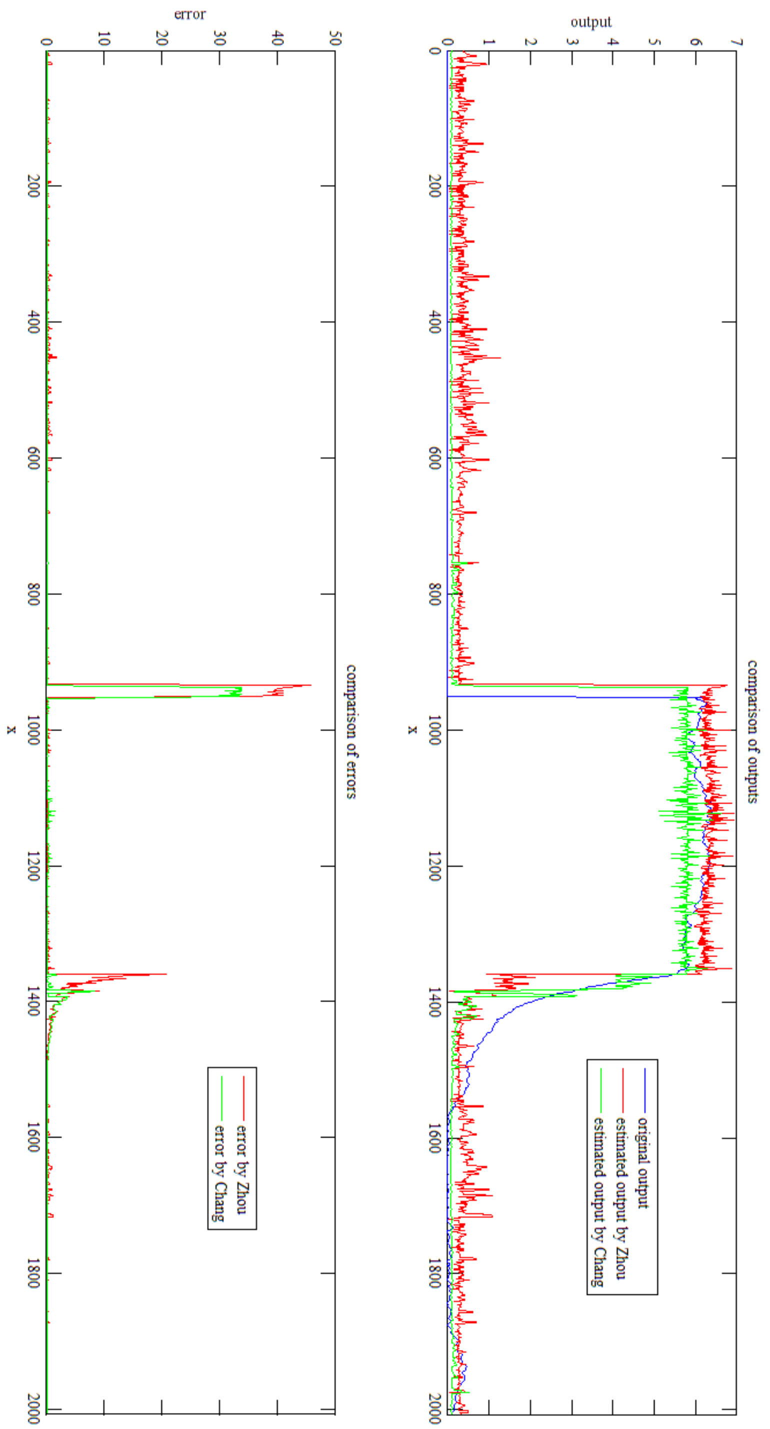


Fig. 6 Comparison on outputs and errors of the estimated and actual results for Case II

Table 6 comparison with present researches

No.	year	description	assumption	MSE (MAE)	AIC_{BRB}	size(training/test)***	NOR
1	2007[32]	local training	conjunctive	0.4049	6549.5039	500/2008	56
2	2009[42]	online updating	conjunctive	0.7880	7130.1742	800/2008	56
3	2010*[41]	sequential learning	conjunctive	0.0241	N/A	305/17	5
4	2011[11]	adaptive learning	conjunctive	0.3990	6536.7031	500/2008	56
5	2013**[21]	extended BRB	conjunctive	(0.2169)	N/A	/2008	900
6	2015[10]	ER rule	disjunctive	0.3709	5945.0172	/2008	15
7	2016[31]	dynamic rule adjustment	conjunctive	0.5040	6212.4300	900/2008	14
				0.4450	6011.8564	900/2008	6
8	**2017[9]**	**bi-level**	**disjunctive**	**0.2917**	**5626.4019**	**500/2008**	**5**
9	2018[38]	JOPS	conjunctive	0.3998(0.1738)	5918.4498	900/2008	6
			disjunctive	0.3974(0.1712)	5871.1991	900/2008	**3**
10	**This study**	**AIC**	**conjunctive**	**0.3411(0.2339)**	**5790.5986**	**500/2008**	**8**

* In [41], BRB is dynamically updated to handle specific inputs. Moreover, its testing dataset is also very small. Therefore, it is also incomplete and thus is not used for comparison.

** In [21], the MSE by the extended BRB is not given. In [21] [38], the MAE is used as the optimization objective instead of MSE and AIC_{BRB} of this study.

*** In [11] [21] [32] [38] [41] [42], the training dataset is selected from specific time period while it is randomly selected in [9] [31] and this study.

Remark 8: The identified optimization path can be recognized as a pareto frontier comprised of multiple optimal solutions. With an identified optimization path, if the weights of different attributes are taken into consideration or an integrated objective AIC_{BRB} is used as in this study, an optimal solution can be identified. For this case, it is found that more optimal solutions are accumulated among the *FlowDiff* attribute. Therefore, for future modeling, *FlowDiff* should be modeled more discretely (with more referenced values).

Remark 9: It should be noted that the comparison of performances of approaches proposed in previous studies should not be directly recognized as their incompetence or superiority. For [11] [32], only parameter learning is considered. For [42], it is designed for online systems and its testing dataset is limited. For [38], its optimize objective is generalization error and MAE instead of AIC_{BRB} and MSE as in this study. Thus, the result of this study is inferior to [38] in comparison of MAE while an improved MSE and AIC_{BRB} are achieved by this study compared with [38]. For [9], it is only applicable for disjunctive BRB (its model and algorithm can not be transferred to conjunctive BRB). Therefore, to make a fair comparison between different models and approaches should have a clearer understanding of the application conditions and restrictions of the respective models and approaches.

6.2.3 Validation on the optimization path search strategy

Since DE is applied as the optimization engine (and also training dataset is randomly selected), this test is conducted in 30 runs. Table 7 shows the optimization process for the 27th , 28th, and 29th runs. An optimal solution is identified as it is with the smallest AIC_{BRB} in comparison with its neighbor solutions. Therefore, Table 7 shows both the "present" solutions (which denote the identified optimal solution) and the solutions to "search" (which denotes the next possible optimal solution(s)). In order to identify the next optimal solution, each attribute is tested to be with one new referenced value.

Table 7 optimization process for the 27th, 28th, and 29th runs

No.		27th run		28th run		29th run	
		optimal solution	AIC_{BRB}	optimal solution	AIC_{BRB}	optimal solution	AIC_{BRB}
1	present	(2, 2)	6220.11	(2, 2)	6186.83	(2, 2)	6181.95
	search	(3,2)/(2,3)	5903.68/6207.20	(3,2)/(2,3)	5961.93/6233.89	(3,2)/(2,3)	5905.57/6209.78
2	present	**(3,2)**	**5903.68**	(3,2)	5961.93	(3,2)	5905.57
	search	(4,2)/(3,3)	5916.06/5933.39	(4,2)/(3,3)	5973.68/5957.75	(4,2)/(3,3)	5790.60/5920.43
3	present			**(3,3)**	**5957.75**	**(4,2)**	**5790.60**
	search			(4,3)/(3,4)	5977.02/6159.8	(4,3)/(5,2)	6013.46/5955.22

Note: (2, 2) denotes two referenced values for the first attribute, *FlowDiff*, and two referenced values for the second attribute, *PressureDiff*.

For each optimization step of each run, an optimal solution is derived which further constitutes a part of an optimization path. In other words, an optimization path is comprised of optimal solutions with given number of dereferenced values for *FlowDiff* and *PressureDiff*.

Table 8 further calculates the statistics of the optimal solutions produced in 30 runs. It shows that (4, 2) is identified as the optimal solution with the highest probability of 40% (12 out of 30 runs), which is consistent with the results from 30 runs: (4, 2) also proves to be the smallest AIC_{BRB} of 5790.5986 (in fact, also the smallest MSE).

Table 8 Statistics on the solutions for Case II

	solution information			number and percentage in 30 runs	
	optimal solution	AIC_{BRB}	MSE	number	percentage
1	(3, 2)	5877.4246	0.3882	7	23.33%
2	(3, 3)	5918.0496	0.3904	1	3.33%
3	(4, 2)	5790.5986	0.3411	12	40.00%
4	(5, 2)	5835.4561	0.3485	6	20.00%
5	(6, 2)	5844.5070	0.3418	4	13.33%

Note: (3, 2) denotes three referenced values for the first attribute, *FlowDiff*, and two referenced values for the second attribute, *PressureDiff*.

To further validate on strategy on selecting the first optimal solution as the final solution, the solution space is further extended to (6, 6).

Fig. 7 gives the results for the 27th run, 28th run, 29th run and the comprised of the optimal solutions from all 30 runs. Further detailed statistics are given in Appendix D.

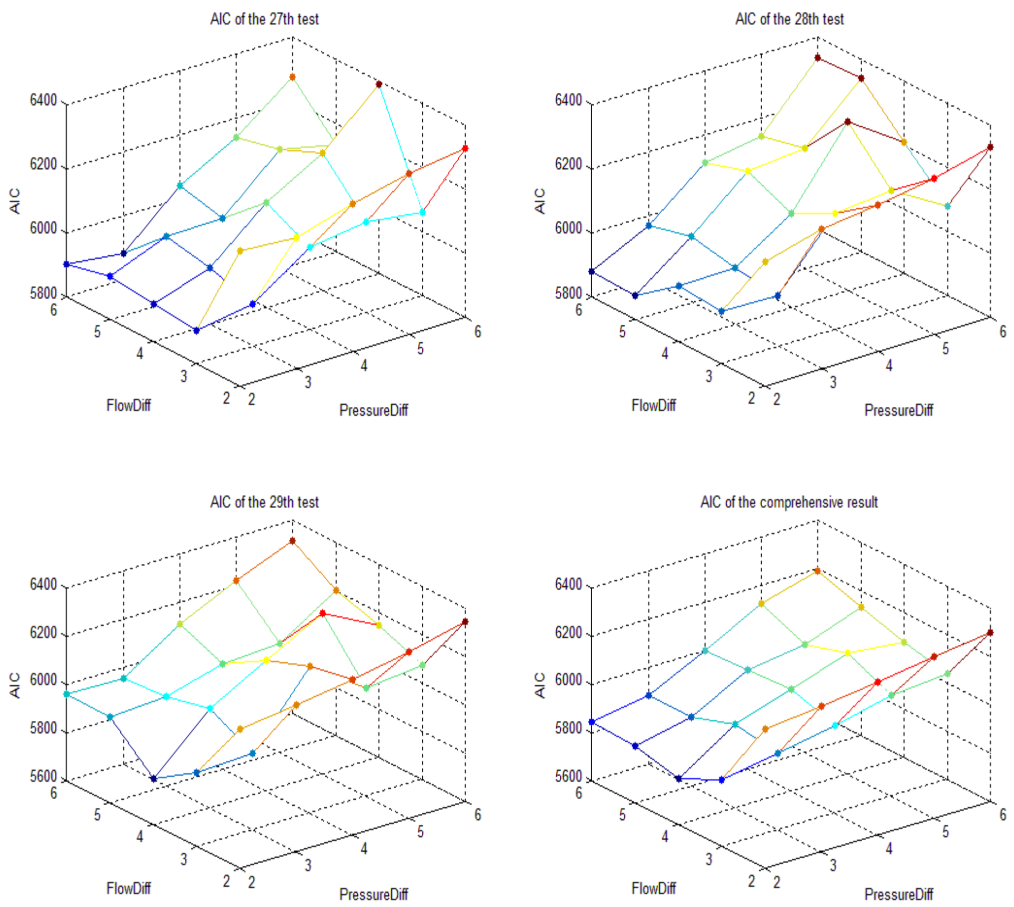


Fig. 7 exploration within the solution space (6, 6) for four examples for Case II

First, under most conditions, the first optimal solution is the local optimal solution in an extended space. Even if it is not, it is rather close. For example, in the 27th run, (3, 2) is identified with the optimal solution with $AIC_{BRB} = 5903.68$ while the local optimal solution (within the solution space of (6, 6)) is (6, 3) with $AIC_{BRB} = 5884.34$. Comparatively, the improvement is only 0.32%. But the computational cost is far bigger. Therefore, although it could be argued that an even superior solution with a smaller AIC_{BRB} could be obtained by devoting more computational resources, it could also be far more expensive and time consuming. Therefore, to identify the first local extreme solution as the final solution is more practical and efficient from an engineering perspective.

Moreover, by exploring the optimization paths identified in 30 runs, the solutions are accumulated along the first attribute, *FlowDiff*, namely (3, 2), (4, 2), (5, 2) and (6, 2), which make up 96.67% (29 out of 30 runs) of the entire solution space. And it leads to the smallest $AIC_{BRB} = 5790.5986$. In other words, the optimization path helps reveals the hidden correlation between the attributes and system's behavior (pipeline leakage in this study). For future modeling on the pipeline leak detection case, more resources should be allocated to modeling *FlowDiff*. To be more specific, *FlowDiff* should be modeled with more referenced values in comparison with *PressureDiff*. By doing so, it could save more modeling resources and computational cost.

Remark 10: There are two possible reasons why (4, 2) produces the smallest MSE other than (5, 2) and (6, 2). First, more referenced values may have resulted in over-fitting [31] which could adversely increase the MSE. Second, DE is set to stop at the 500th generations. This may be insufficient for (5, 2) or (6, 2) since there were more parameters. In other words, a smaller MSE could be achieved by consuming more computational time and resources. However, the optimization algorithm can not go on forever, a stop criterion must be applied.

Remark 11: The optimization path can help reveal hidden correlation between the attributes of complex systems with its behavior. Take this pipeline leak detection case as an example. Upon identifying the optimization path, it is found that the attribute *FlowDiff* have a stronger correlation with the pipeline leakage. Thus, it should be modeled with more discretion (more referenced values) in order to achieve better modeling performance as well as saving modeling resources and computation time.

Remark 12: As shown by the optimization process, the optimal solution can be identified within limited steps. This is because the first optimal solution is recognized as the final solution. As discussed in Remark 6, the computational complexity is linear to the number of attributes by adopting such strategy. Moreover, it is also computational efficient from a practical perspective.

7. Conclusion

Driven by two motives, to jointly optimize the structure and parameters for the conjunctive BRB and search for the optimization path, this study proposes a new conjunctive BRB learning approach.

For jointly optimizing the structure and parameter for the conjunctive BRB, a new BRB learning approach is proposed with a bi-level optimization model with AIC_{BRB} as the integrated objective and a corresponding bi-level optimization algorithm.

For optimization path search, it is comprised of multiple optimal solutions. An identified optimization path can be used to determine the final optimal solution as well as reveal hidden correlation between the attributes of complex systems with its behavior.

Two cases are studied to validate the efficiency of the proposed BRB learning approach. In the numerical case, it is found that the optimal BRB the modeling accuracy and the modeling complexity are both increased to achieve a smaller AIC_{BRB}. In the practical case, in comparison with previous studies, the MSE is reduced and even fewer rules/parameters are needed which leads to a far smaller AIC_{BRB}.

As for the optimization path search strategy, it is found that the majority of the optimal solutions (29 out of 30 runs) are accumulated in the "*FlowDiff*" attribute. Therefore, *FlowDiff* should be designed with more referenced values to improve the efficiency for future modeling and optimization.

It should again be emphasized that the BRB learning approach proposed in this study is designed specifically for conjunctive BRB because different types of BRBs are constructed differently and thus requires different learning approach. For future research, the application conditions and restraints for different types of BRB should be determined based on characteristics of the research subject.

Acknowledgements

This work was supported by the National Science Foundation of China under Grants 71601180, 71671186, 71401167, 61370031, 61374138, 61773388, 61751304, and 61702142. The NSFC-Zhejiang Joint Fund for the Integration of Industrialization and Informatization (U1709215), the Post-Doctorial Science Foundation under Grants 2015M570847 and 2016T90938, and the Natural Science Foundation of Hainan Province under Grant 617120. This work was also supported by the MOE (Ministry of Education in China) Liberal Arts and Social Sciences Foundation (Nos. 17YJCZH157) and the Pengcheng Scholar Funded Scheme.

References

[1] H. Akaike, Information theory and an extension of the maximum likelihood principle, in Proc. 2nd Int. Symp. Inform. Theory, 1971, pp. 267-281.

[2] H. Akaike, Akaike’s Information Criterion, International Encyclopedia of Statistical Science, Springer Berlin Heidelberg, 2011.

[3] D. R. Ball, A. Deshmukh, N. Kapadia, An options-based approach to coordinating distributed decision systems, Eur. J. Oper. Res. 240 (2015) 706-717.

[4] H. Bozdogan, Model selection and Akaike's information criterion (AIC): The general theory and its analytical extensions, Psychometrika 52(3) (1987) 345-370.

[5] K. P. Burnham, D. P. Anderson, Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach, London: Springer, 2002.

[6] L. L. Chang, L. Y. Wang, G. Liu, W. Wang, Akaike Information Criterion-based objective for Belief Rule Base optimization, in Proc. 8th Int. Conf. Intell. Human-Mach. Syst. Cybern., 2016, pp. 545-549.

[7] L. L. Chang, Z. J. Zhou, Y. W. Chen, T. J. Liao, Y. Hu and L. H. Yang, Belief Rule Base Structure and Parameter Joint Optimization Under Disjunctive Assumption for Nonlinear Complex System Modeling, IEEE T. Syst. Man Cybern. Systs., doi: 10.1109/TSMC.2017.2678607.

[8] 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.

[9] L. L. Chang, Z. J. Zhou, Y. You, L. H. Yang, Z. G. Zhou, Belief rule based expert system for classification problems with new rule activation and weight calculation procedures, Inf. Sci. 335 (2016) 75-91.

[10] Y. Chen, Y. W. Chen, X. B. Xu, C. C. Pan, J. B. Yang, G. K. Yang, A data-driven approximate casual inference model using the evidential reasoning rule, Knowl. Based Syst. 88 (2015) 264-272.

[11] Y. W. Chen, J. B. Yang, D. L. Xu, Z. J. Zhou, D. W. Tang, Inference analysis and adaptive training for belief rule based system, Expert Syst. Appl. 38 (2011) 12845-12860.

[12] B. Colson, P. Marcotte, G. Savard, An overview of bilevel optimization, Ann. Oper. Res. 153 (2007) 235-256.

[13] D. L. Donoho, High-dimensional data analysis: The curses and blessings of dimensionality. AMS Math Challenges Lecture, 1 (2000): 32.

[14] C. Fu, S. L. Yang, An attribute weight based feedback model for multiple attributive group decision analysis problems with group consensus requirements in evidential reasoning context, Eur. J. Oper. Res. 212 (2011) 179-189.

[15] J. Gordon, E. H. Shorliffe, A method for managing evidential reasoning in a hierarchical hypothesis space, Artif. Intell. 26(3) (1985) 323-357.

[16] J. F. Gu, X. J. Tang, Meta-synthesis approach to complex system modeling, Eur. J. Oper. Res. 166 (2006) 597-614.

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

[18] R. M. Kovacevic, G. C. Pflug, Electricity swing option pricing by stochastic bilevel optimization: a survey and new approaches, Eur. J. Oper. Res. 237(2) (2014) 389-403.

[19] G. L. Li, Z. J. Zhou, C. H. Hu, L. L. Chang, Z. G. Zhou, F. J. Zhao, A new safety assessment model for complex system based on the conditional generalized minimum variance and the belief rule base, Safety Sci., 93 (2007) 108-120.

[20] A. van der Linde, DIC in variable selection, Stat. Neerlandica 59(1) (2005) 45-56.

[21] 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.

[22] L. Montiel, R. Dimitrakopoulus, Optimizing mining complexes with multiple processing and transportation alternatives: An uncertainty-based approach, Eur. J. Oper. Res. 247 (2015) 166-178.

[23] K. Price, Differential Evolution vs. the functions of the 2nd ICEO, in Proc. 1997 IEEE Int. Conf. Evolut. Comput., 1997, pp. 153-157.

[24] M. Ryoke, Y. Nakamori, Agent-based approach to complex system modeling, Eur. J. Oper. Res. 166 (2005) 717-725.

[25] G. E. Schwarz, Estimating the dimension of a model, Ann. Stat. 6(2) (1978) 461-464.

[26] A. Sinha, P. Malo, K. Deb, Evolutionary algorithm for bilevel optimization using approximations of

the lower level optimal solution mapping, Eur. J. Oper. Res. 257 (2017) 395-411.

[27] R. Storn, K. Price, Differential Evolution - A Simple and Efficient Heuristic for Global Optimization over Continuous Spaces, J. Global Optim. 11(4) (1997) 341-359.

[28] E. E. Savan, J. B.Yang, D. L. Xu, Y. W.Chen, A genetic algorithm search heuristic for belief rule base model-structure validation, in Proc. 2013 IEEE Int. Conf. Syst., Man, Cybern., 2013, pp. 1373-1378.

[29] W. B. Wang, A simulation-based multivariate Bayesian control chart for real time condition-based maintenance of complex systems, Eur. J. Oper. Res. 218 (2012) 726-734.

[30] Y. M. Wang, J. B. Yang, D. L. Xu, Environmental impact assessment using the evidential reasoning approach, Eur. J. Oper. Res. 174 (2006) 1885-1913..

[31] 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.

[32] D. L. Xu, J. 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 (2007) 103-113.

[33] D. L. Xu, S. L. Yang, Integrated efficiency and trade-off analysis using DEA-oriented interactive minimax reference point approach, Comput. Oper. Res. 39 (2012) 1062-1073.

[34] J.B. Yang, Rule and utility based evidential reasoning approach for multi-attribute decision analysis under uncertainties, Eur. J. Oper. Res. 131 (2001) 31-61.

[35] J. B. Yang, J. Liu, D. L. Xu, J. Wang, Y. M. Wang, Optimization models for training belief-rule-based systems, IEEE T. Syst. Man Cybern. A-Systs. Hum. 37 (2007) 569-585.

[36] J. B. Yang, Y. M. Wang, D. L. Xu, K. S. Chin, L. Chatton, Belief rule-based methodology for mapping consumer preference and setting product targets, Expert Syst. Appl. 39 (2012) 4749-4759.

[37] J. B. Yang, D. L. Xu, Evidential reasoning rule for evidence combination, Artif. Intell. 205 (2013) 1-29.

[38] L. H. Yang, Y. M. Wang, J. Liu, L. Martines, A joint optimization method on parameter and structure for belief-rule-based systems, Knowl. Based Syst. 142 (2018) 220-240.

[39] Z. G. Zhou, F. Liu, L. C. Jiao, et al. A bi-level belief rule based decision support system for diagnosis of lymph node metastasis in gastric cancer, Knowl. Based Syst. 54(4) (2013)128-136.

[40] Z. G. Zhou, F. L. Liu, L. L. Li, et al. A cooperative belief rule based decision support system for lymph node metastasis diagnosis in gastric cancer. Knowl. Based Syst. 85 (2015) 62-70.

[41] 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.

[42] 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.

[43] Z. J. Zhou, C. H. Hu, J. B. Yang, D. L. Xu, D. H. Zhou, New model for system behavior prediction based on belief rule based systems, Inf. Sci. 180 (24) (2010) 4834-4864.

Appendix A: the difference between conjunctive and disjunctive BRB optimization

The main difference between conjunctive and disjunctive BRB optimization is that the research subject is different. To be more specific: the number of referenced values for the attributes determines the size of a conjunctive BRB while the size of a disjunctive BRB determines the number of referenced values for the attributes. Thus, the attributes are separately optimized for the conjunctive BRB while they are jointly optimized for disjunctive BRB. Consequently, an optimization path could be identified for conjunctive BRB while it is inapplicable for disjunctive BRB.

We would explain the above difference in details as follows.

- The difference between the conjunctive and disjunctive BRBs.

Both conjunctive and disjunctive BRBs are comprised of multiple rules in the same belief structure. A conjunctive rule is as follows:

$$\begin{aligned} R_k: \text{if} \quad & \left(A_1 \text{ is } x_1^k\right) \wedge \left(A_2 \text{ is } x_2^k\right) \wedge \cdots \wedge \left(A_M \text{ is } x_M^k\right), \\ & \text{then } \{(D_1, \beta_{1,k}), \cdots, (D_S, \beta_{S,k})\} \\ & \text{with rule weright } \theta_k \end{aligned} \tag{A.1}$$

A disjunctive rule is as follows:

$$\begin{aligned} R_k: \text{if} \quad & \left(A_1 \text{ is } x_1^k\right) \vee \left(A_2 \text{ is } x_2^k\right) \vee \cdots \vee \left(A_M \text{ is } x_M^k\right), \\ & \text{then } \{(D_1, \beta_{1,k}), \cdots, (D_S, \beta_{S,k})\} \\ & \text{with rule weright } \theta_k \end{aligned} \tag{A.2}$$

Although the basic form of the conjunctive and disjunctive BRBs seems rather similar, they are quite different in construction. As shown in Fig. A.1(a/b), the conjunctive BRB requires covering each possible combination of the referenced values for the attributes while the disjunctive attributes requires only covering all of the reference values for the attributes.

Note that disjunctive BRB is also “complete” (can handle all inputs regardless of its values) by projecting its reference values to different dimensions, as demonstrated by Fig. A.1(d). ([9] gives the explanation on why disjunctive BRB is still complete.)

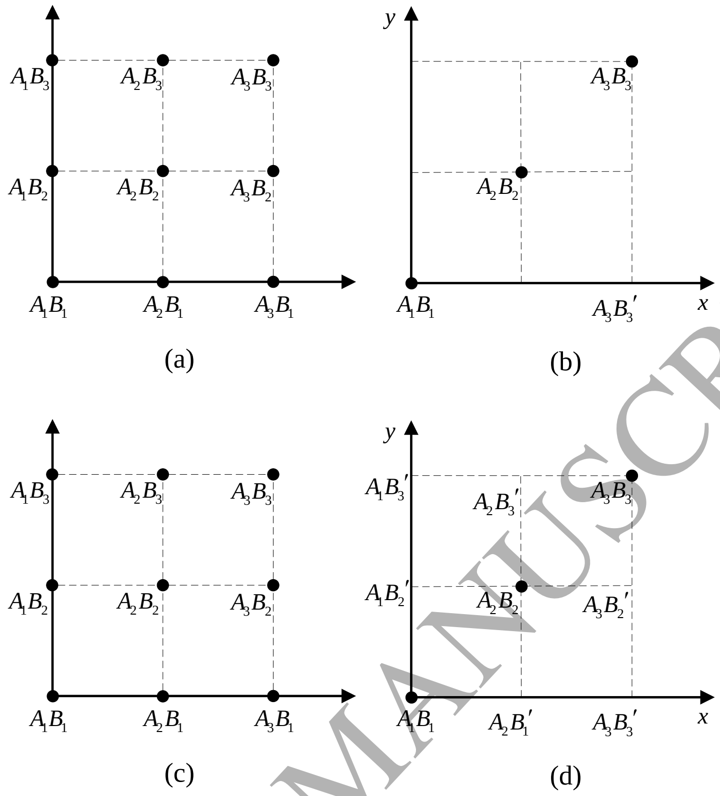


Fig. A.1 conjunctive and disjunctive BRB construction

- The difference between the conjunctive and disjunctive BRB optimization.

For a conjunctive BRB, the referenced values of the attributes are determined first. Afterwards, the structure of a conjunctive BRB is determined. Suppose that a complex system is with 3 attributes, namely x_1, x_2 and x_3. The three attributes are assumed to be with 3, 4 and 5 referenced values (either by experts' knowledge or optimization), then the conjunctive BRB would be 120(=3×4×5) rules. With the determined belief structure, a conjunctive BRB can be completed by adding the beliefs for the scales in the conclusion part of all rules (either by experts' knowledge or optimization).

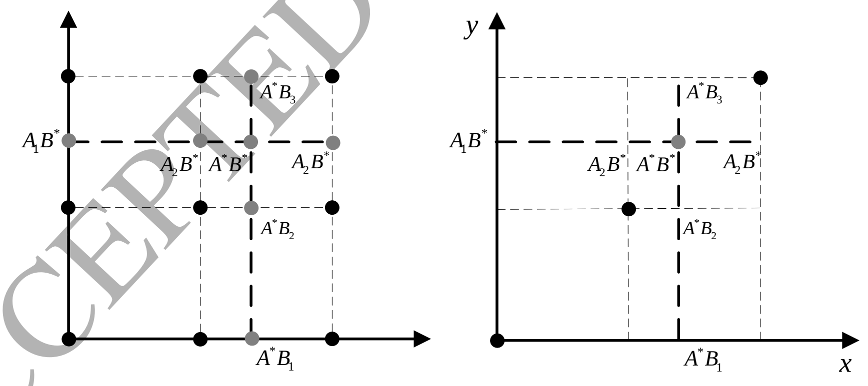
However, for a disjunctive BRB, this process is different. For a disjunctive BRB, normally its size is determined first. Still take the above example and supposedly a disjunctive BRB is applied. We would have to first determine the size of the disjunctive BRB (supposedly 5 rules). Afterwards, each of the three attributes would be assumed to be with no more than 5 referenced values (normally each attribute would be assumed with 5 referenced values for convenience in BRB construction). Thus, the belief structure of the disjunctive BRB is determined. Similarly, a disjunctive BRB can be further completed by adding the beliefs for the scales in the conclusion part of all rules (either by experts' knowledge or optimization).

It is obvious that the construction process for different types of BRB is different. It also results in the difference in optimization.

For conjunctive BRB, its optimization process starts from each single attribute. By adding a new referenced value for an attribute, multiple rules are automatically added to a conjunctive BRB. For example, as in Fig. A.2(a), if A^* and B^* are added to Attribute A and B, respectively. Then 7 new rules would be added to this conjunctive BRB, namely, A_1B^*, A_2B^*, A^*B^*, A_3B^*, A^*B_1, A^*B_2, and A^*B_3. For conjunctive BRB optimization, each attribute is separately optimized, and an optimal solution is derived upon iteratively comparing the optimization performance of different attributes. **To summarize, for conjunctive BRB joint optimization, the attributes are separately optimized**.

For disjunctive BRB, this process is different. By adding a new rule, there would be one new referenced value added to each attribute (through projecting the referenced value of the new rule to each attribute), as shown in Fig. A. 2(b). When optimizing, an optimal solution is derived upon comparing the original BRB with the one with a new rule. In other words, **for disjunctive BRB joint optimization, the attributes are simultaneously optimized**.

This is the fundamental difference between the two types of BRB regarding on optimization. Moreover, an optimization path can be identified through conjunctive BRB optimization while there would not be an optimization path by disjunctive BRB optimization.



(a) (b)

Fig. A.2 conjunctive and disjunctive BRB optimization by introducing new rules

- The optimization search for conjunctive BRB.

Moreover, an optimization path can be identified through conjunctive BRB optimization while there would not be an optimization path by disjunctive BRB optimization. Next, we would further explain on the optimization search for conjunctive BRB.

The optimization path search strategy is illustrated by Fig. 4 in Section 5.3.2. The optimization path search strategy is an iterative process. In each iteration, the performance of BRB with a newly added referenced value for an attribute is compared with that of the previous BRB.

Unlike conventional structure learning which intends to downsize BRB or parameter learning which aims to improving modeling accuracy, an optimization path is a scheme comprised of selected referenced values for the attributes with the optimal performance.

An optimization path can also be recognized as a pareto frontier which is comprised of multiple optimal solutions. With an integrated criterion, an optimal solution can be identified. Note that although AIC_{BRB} is used as the integrated objective in this study, decision makers' preference and other criterions can also be taken into consideration (could be in the form of weights for attributes) and a different optimal solution with consideration of preference can be thus determined.

Moreover, if multiple optimal solutions are accumulated among one attribute or shows a pattern as a "path", then a hidden correlation between the attributes of the complex system and its behavior can thus be obtained. For future modeling and optimization, more resources should be devoted to such attribute and optimization path to save limited resources and computational time.

To summarize, the optimization path serves for two purposes. The first one is to help identify the optimal solution. With different integrated criterion and/or decision makers' preference, different optimal solution can be identified. The second one is to serve as guidance for future modeling and optimization. With an identified optimization path, the hidden correlations between the attributes and system behavior can be revealed. It can be used as guidance for future modeling and optimization in order to save limited resources and computational time.

- The selection of conjunctive BRB or disjunctive BRB

Note that which type of BRB should be constructed is primarily determined by the nature of the complex system. In other words, a conjunctive BRB is applicable if the behavior of that system is the conjunctive result of its attributes, and vice versa. However, sometimes the nature of the system is unclear. Moreover, the preference or judgment of experts has a great influence on the selection of a conjunctive/disjunctive BRB. At such time, whether a conjunctive or a disjunctive BRB should be adopted is mostly determined by experts based on their preference, especially when the mechanism of a complex system is unclear.

Appendix B: the evolutionary operators of DE as the optimization engine

Coding

All parameters are initiated as individuals comprised of decimal coded genes and do not need to be decoded. For the *i*th individual, the genes are coded as in Fig. B.1.

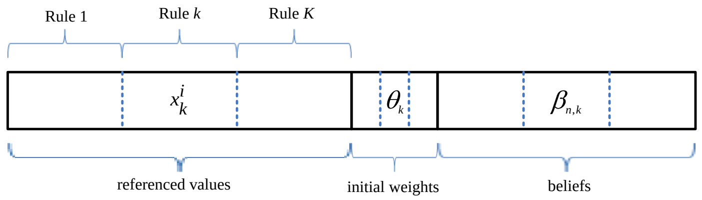


Fig. B.1 the gene coding for the *i*th individual

Crossover

The crossover strategy states that the *j*th gene, $v'_{i,j}$, of a temporary individual, v, is selected by the probability of CR (or the *j*th gene, $z'_{i,j}$, of the current individual, z, is selected by the probability of $1-CR$) as the *j*th gene, $u'_{i,j}$, of the final individual, u, as shown in Eq. (B.1),

$$u'_{i,j} = \begin{cases} v'_{i,j} & \text{if } (rand \le CR) \text{ or } (j = sn) \\ z'_{i,j} & \text{otherwise} \end{cases} \tag{B.1}$$

where $CR = 0.9$ is the crossover operator and $sn \in [1,2,\cdots,n]$ is a random integer which is generated with each new individual.

Mutation

The *i*th individual in the new generation, v'_i, can be obtained using Eq. (B.2),

$$v'_i = z_{r1} + F * (z_{r2} - z_{r3}) \tag{B.2}$$

where z_{r1}, z_{r2} and z_{r3} are three random individuals and $r1 \ne r2 \ne r3$. $F = 0.5$ is the mutation operator.

Selection

For the *i*th individual u_i^t, it enters the new generation when its fitness function get a higher rated value, as indicated in Eq. (B.3),

$$z_i^{t+1} = \begin{cases} u_i^t & \text{if } \quad f(u_i^t) \le f(z_i^t) \\ z_i^t & \text{otherwise} \end{cases} \tag{B.3}$$

where $f(\bullet)$ is the fitness function which in this study is the MSE.

Appendix C: the rule activation, matching degree calculation and weight calculation

Consider that the targeted problem is with two attributes, namely A and B. Each of the two attributes has three referenced values. In order to cover all the possible conditions, a traversal BRB system which consists of 9 (=3×3) rules is constructed. In this condition, suppose that there is one input, namely *I*, as shown in Fig. C.1, and then four adjacent points/rules would be activated by each input data, namely A_2B_3, A_3B_3, A_2B_2 and A_3B_2.

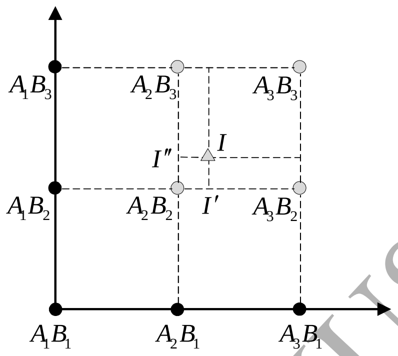


Fig. C.1 conventional rule activation

For the input, $(x_1^*,\cdots,x_m^*,\cdots,x_M^*)$, the matching degree of the *m*th attribute in the *j*th rule is calculated by Eq. (C.1),

$$\varphi(x_m^*, x_m^j) = \begin{cases} \dfrac{x_m^{k+1} - x_m^*}{x_m^{k+1} - x_m^k} & j = k(x_m^k \le x_m^* \le x_m^{k+1}) \\ \dfrac{x_m^* - x_m^k}{x_m^{k+1} - x_m^k} & j = k+1 \\ 0 & j = 1, 2, \ldots |x_i|,\ j \ne k, k+1 \end{cases} \tag{C.1}$$

where x_m^* denotes the value of the *m*th attribute in the input data. x_m^{k+1} and x_m^k denote the values of the *m*th attribute in two adjacent activated rules (the *k*th and the (*k*+1)th rule).

The integrated matching degree for the *m*th attribute in the *j*th rule is calculated by Eq. (C.2),

$$\alpha_m^j = \frac{\varphi(x_m^*, x_m^j)c_m}{\sum \varphi(x_m^*, x_m^j)} \tag{C.2}$$

where c_m denotes the confidence of the *m*th attribute being assessed as x_m^*. If there is no incomplete information ($\sum \varphi(x_m^*, x_m^j) = 1$) and the value of the *m*th attribute is of 100% confidence ($c_m = 1$), then,

$$\alpha_m^j = \varphi(x_m^*, x_m^j) \tag{C.3}$$

The activated weight for the kth rule is calculated by Eq. (C.4),

$$w_k = \frac{\theta_k \prod_{m=1}^{M} \alpha_m^k}{\sum_{k=1}^{K} \theta_k \prod_{m=1}^{M} \alpha_m^k} \tag{C.4}$$

Appendix D: statistics for the 27th, 28th, 29th tests and the composite result

Table D.1 statistics for the 27th test

First rules	2	2	3	2	4	3	5	2	3
Second rules	2	3	2	4	2	3	2	5	4
No. rules	4	6	6	8	8	9	10	10	12
No. parameters	24	37	37	50	50	56	63	63	75
MSE	0.5926	0.5667	0.4001	0.5856	0.3939	0.3963	0.3900	0.5948	0.4372
AIC	6220.11	6207.20	5903.68	6261.81	5916.06	5933.39	5933.57	6301.32	6057.12
First rules	4	2	6	5	3	4	3	6	5
Second rules	3	6	2	3	5	4	6	3	4
No. rules	12	12	12	15	15	16	18	18	20
No. parameters	75	76	76	94	94	100	113	113	125
MSE	0.3983	0.5960	0.3659	0.3949	0.4300	0.4461	0.4011	0.3287	0.3692
AIC	5975.88	6329.09	5903.86	6006.46	6080.59	6124.63	6058.00	5884.34	6009.61
First rules	4	4	6	5	5	6	6		
Second rules	5	6	4	5	6	5	6		
No. rules	20	24	24	25	30	30	36		
No. parameters	125	150	150	158	187	187	224		
MSE	0.4741	0.5392	0.3611	0.4123	0.3683	0.3720	0.3985		
AIC	6227.60	6389.84	6040.30	6171.88	6131.61	6140.40	6274.30		

Table D.2 statistics for the 28th test

First rules	2	2	3	2	4	3	5	2	3
Second rules	2	3	2	4	2	3	2	5	4
No. rules	4	6	6	8	8	9	10	10	12
No. parameters	24	37	37	50	50	56	63	63	75
MSE	0.5704	0.5843	0.4277	0.5813	0.4208	0.4075	0.3644	0.5835	0.4919
AIC	6186.83	6233.89	5961.93	6255.37	5973.68	5957.75	5874.34	6284.71	6159.80
First rules	4	2	6	5	3	4	3	6	5

Second rules	3	6	2	3	5	4	6	3	4
No. rules	12	12	12	15	15	16	18	18	20
No. parameters	75	76	76	94	94	100	113	113	125
MSE	0.3988	0.5981	0.3563	0.3944	0.4816	0.4297	0.4099	0.3632	0.4367
AIC	5977.02	6332.25	5880.79	6005.23	6179.31	6092.04	6076.90	5971.41	6156.14
First rules	4	4	6	5	5	6	6		
Second rules	5	6	4	5	6	5	6		
No. rules	20	24	24	25	30	30	36		
No. parameters	125	150	150	158	187	187	224		
MSE	0.5305	0.4379	0.3921	0.4130	0.4681	0.3723	0.4277		
AIC	6325.63	6208.36	6112.15	6173.41	6340.67	6140.97	6335.96		

Table D.3 statistics for the 29th test

First rules	2	2	3	2	4	3	5	2	3
Second rules	2	3	2	4	2	3	2	5	4
No. rules	4	6	6	8	8	9	10	10	12
No. parameters	24	37	37	50	50	56	63	63	75
MSE	0.5672	0.5684	0.4009	0.5755	0.3411	0.3904	0.3998	0.5874	0.5219
AIC	6181.95	6209.78	5905.57	6246.64	5790.60	5920.43	5955.22	6290.48	6211.43
First rules	4	2	6	5	3	4	3	6	5
Second rules	3	6	2	3	5	4	6	3	4
No. rules	12	12	12	15	15	16	18	18	20
No. parameters	75	76	76	94	94	100	113	113	125
MSE	0.4159	0.6084	0.3909	0.3787	0.4146	0.4557	0.4078	0.3574	0.3797
AIC	6013.46	6347.02	5961.41	5969.87	6048.85	6143.21	6072.31	5957.29	6034.20
First rules	4	4	6	5	5	6	6		
Second rules	5	6	4	5	6	5	6		
No. rules	20	24	24	25	30	30	36		
No. parameters	125	150	150	158	187	187	224		
MSE	0.4970	0.4091	0.3914	0.3601	0.4006	0.4090	0.4194		
AIC	6268.87	6149.10	6110.54	6054.03	6204.79	6222.91	6318.74		

Table D.4 statistics for the composite result

First rules	2	2	3	2	4	3	5	2	3
Second rules	2	3	2	4	2	3	2	5	4

No. rules	4	6	6	8	8	9	10	10	12
No. parameters	24	37	37	50	50	56	63	63	75
MSE	0.5662	0.5667	0.3882	0.5693	0.3411	0.3894	0.3485	0.5746	0.3925
AIC	6180.37	6207.14	5877.43	6237.11	5790.60	5918.05	5835.46	6271.21	5963.00
First rules	4	2	6	5	3	4	3	6	5
Second rules	3	6	2	3	5	4	6	3	4
No. rules	12	12	12	15	15	16	18	18	20
No. parameters	75	76	76	94	94	100	113	113	125
MSE	0.3853	0.3882	0.3418	0.3446	0.4015	0.3969	0.3921	0.3287	0.3692
AIC	5946.86	5955.43	5844.51	5887.59	6020.79	6022.77	6038.23	5884.34	6009.61
First rules	4	4	6	5	5	6	6		
Second rules	5	6	4	5	6	5	6		
No. rules	20	24	24	25	30	30	36		
No. parameters	125	150	150	158	187	187	224		
MSE	0.4116	0.3774	0.3458	0.3581	0.3683	0.3669	0.3617		
AIC	6104.50	6078.96	6002.53	6045.20	6131.61	6128.29	6189.73		