Enhancing Extended Belief Rule-Based Systems for Classification Problems Using Decomposition Strategy and Overlap Function

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

[a]Decision Sciences Institute, Fuzhou University, Fuzhou, PR China

[b]Key Laboratory of Spatial Data Mining & Information Sharing of Ministry of Education, Fuzhou University, Fuzhou, PR China

[c]School of Computing, Ulster University, Northern Ireland, UK

[d]Department of Computer Science, University of Jaén, Jaén, Spain

*The corresponding author. E-mail: msymwang@hotmail.com

Abstract: Multi-class and multi-attribute are two important features of classification problems and have different effects on the requirements and performance of the classifier. *Decomposition strategy* and *overlap function* are two effective ways to enhance the performance of classifiers, because the former decomposes a complex multi-class problem into multiple simple sub-problems; the latter uses various functions to specify the conjunctive relationship of input variables in a multi-attribute problem. *Extended belief rule-based system* (EBRBS) is an advanced rule-based system that has been widely used in classification problems. In order to apply decomposition strategies and overlap functions to enhance the performance of EBRBSs, the present work focuses on the investigative research and comparative evaluation of the commonly used one-versus-one (OVO) decomposition strategy and five common overlap functions to improve the performance of EBRBSs on multi-class and multi-attribute problems. More specifically, three typical kinds of EBRBSs, namely original EBRBS (O-EBRBS), EBRBS with dynamic rule activation (DRA-EBRBS), and a latest EBRBS for big data (Micro-EBRBS), are selected to conduct extensive experimental studies on twenty classification problems. To best of our knowledge, this present work is the first time to provide a meaningful and useful study in revealing the potential capability of the EBRBSs with decomposition strategy and overlap function for multi-class and multi-attribute problems. Experimental results demonstrate that the square product overlap function and the OVO strategy can enhance the performance of EBRBSs over others for twenty classification problems.

Keywords: Extended belief rule-based system; Overlap function; Decomposition strategy; One-versus-one; Classification

1. Introduction

Classification is a fundamental and critical problem to be addressed in various theoretical and practical applications, *e.g.*, image processing [1][2], pattern recognition [3][4], and intelligent transportation [5][6]. Within these real-world problems, two types of problems can be differentiated according to the number of classes in output variables and the number of attributes in input variables: multi-class and multi-attribute problems. In general, a classifier is more difficult to obtain a desired performance for multi-class problems because of the increased complexity in each class boundary, caused by the higher intersection among the data of different classes. Meanwhile, multi-attribute problems are a difficult task for a classifier due to the low variation between data, produced by the conjunctive relationship between various variables existed in multi-attribute problems. Even so, it is inevitable that multi-class and multi-attribute problems have to be considered when a classifier is applied to address classification problems.

To overcome the difficulties caused by multi-class and multi-attribute in a classification problem, *decomposition strategy* and *overlap function* were introduced to enhance the performance of classifiers in many previous studies [7][8] (See further detailed literature reviews in Section 2). Throughout these attempts there is a feasible solution for other up-rising classifiers when addressing classification problems with multiple classes and multiple attributes, even the classifiers have the inherent multi-class and multi-attribute supports. It is worth noting that decomposition strategy can be described as the process of dividing a complex multi-class classification problem into multiple easier-to-solve ones, *i.e.*, the original classification problem is related with three classes $\{D_1, D_2, D_3\}$ and it can be divided into multiple two-class problems, like $\{D_1, D_2\}$, $\{D_1, D_3\}$ and $\{D_2, D_3\}$; and overlap function can be described as a linear or nonlinear function which allows classifiers to obtain the aggregated values with a higher variation for the conjunctive relationship among input variables found in multi-attribute problems, *i.e.*, for a multi-attribute problem with three input variables $\{x_1, x_2, x_3\}$, the conjunctive relationship of these input variables can be wrote as $f(x_1, x_2, x_3) = x_1 \times x_2 \times x_3$ when using product function as an overlap function.

Among various kinds of decomposition strategies, one-versus-one (OVO) decomposition strategy is one of the most common strategies used in multi-class problems to divide an N-class problem into $N \times (N - 1) / 2$ binary sub-problems (each two from N classes). Based upon this viewpoint, when the OVO strategy is applied to enhance a classifier, each binary sub-problem should be used to model an independent binary sub-classifier and the output of all binary sub-classifiers needs to be aggregated to make a final decision for replying any given input data, where some common aggregation methods include voting strategy, weighted voting, WinWV, non-dominance criteria, and learning valued preference for classification. Owing to the OVO decomposition strategy, classifiers usually obtain a better result than addressing multi-class classification problems directly, even so the classifiers have the inherent multi-class support. Furthermore, the OVO decomposition strategy also provides an available framework to enhance classifiers by using ensemble learning [9][10] and parallelization techniques [11][12].

Among various kinds of overlap functions, square product, product, minimum, geometric mean, and sine functions are the widely used overlap functions for classifiers when addressing multi-attribute classification problems in the previous studies [13][14], and all these overlap functions have been proven to be able to effectively model the conjunctive relationship between the input variables of a multi-attribute problem, which are used as the inputs of classifiers, and the classifiers therefore produce different kinds of the aggregated values with different variations made by different numbers of attributes, *i.e.*, the outcomes of square product, product, minimum, geometric mean, and sine functions are sorted in ascending order when all these functions are assumed to have the same input variables within range [0, 1]. As a result, it is evident to offer a panoramic view of the performance of a certain classifier for dealing with multi-attribute classification problems and provide a solution to enhance the classifier.

In recent years, many effective and efficient classifiers [11][15]-[17] were developed for classification problems on the basis of the *extended belief rule-based system* (EBRBS), which was proposed by Liu *et al*. [18] by embedding belief structures into the rule antecedent of belief rules so that both rule antecedent and consequent can represent uncertain information, *e.g*., fuzzy uncertainty, random uncertainty, and incomplete uncertainty. Thus, the classifiers based on the EBRBS have important advances in uncertainty information processing and modeling. However, similar to other methodologies used in developing an advanced classifier [19]-[22], the EBRBS has to encounter the difficulties caused by multi-class and multi-attribute in a classification problem as well. Providing feasible solutions to improve the performance of EBRBS-based classifiers is a critical problem that needed to be solved when addressing multi-class and multi-attribute classification problems.

However, the previous studies of enhancing EBRBS mainly focused on the determination of consistent activation rules [15][23][24] and the reduction of ineffective rules [11][25][26], because the original EBRBS (O-EBRBS for short) lacks of necessary mechanisms to avoid two undesired situations: 1) the set of activated rules is based on the conflicting information to reply input data; 2) the size of an extended belief rule rule (EBRB) would increase unlimitedly. For this purpose, two representative EBRBSs extended from O-EBRBS were proposed by considering dynamic rule adjustment (DRA) method and domain division-based rule reduction method, respectively, and they are therefore called DRA-EBRBS [15] and Micro-EBRBS [11]. On the background of these three EBRBSs, it can be found that they were only based on product function to model the conjunctive relationship in multi-attribute problems and have not been investigated yet for multi-class problems by using decomposition strategies. Hence, four motivations of the present work can be summarized as follows:

(1) O-EBRBS, DRA-EBRBS and Micro-EBRBS are introduced to perform an investigative study and comparative analysis for the intention of distinguishing the performance differences existed in these three representative EBRBSs.

(2) The influence of the OVO decomposition strategy with five aggregation methods on the three EBRBSs is investigated to provide potential solutions when it is required to enhancing EBRBSs for addressing multi-class problems.

(3) The influence of the square product, product, minimum, geometric mean, and sine overlap functions on the three EBRBSs is investigated to provide potential solutions in the case of enhancing EBRBSs for addressing multi-attribute problems.

(4) The last motivation of the proposed work is to place important basis and framework for EBRBSs to be further enhanced through ensemble learning and parallelization techniques in terms of both accuracy and efficiency.

Based on the above-mentioned four motivations, an experimental study considering twenty classification datasets obtained from the KEEL dataset repository [27] is carried out to achieve well-founded conclusions. As suggested in the specialized literature [28], a non-parametric statistical test, namely Aligned Friedman and Holm test, is utilized to further confirm the influence of the OVO decomposition strategy and the five overlap functions on the three EBRBSs for multi-class and multi-attribute classification problems. Correspondingly, the main contributions of this work can be summarized as follows:

(1) A comparison of O-EBRBS, DRA-EBRBS, and Micro-EBRBS is carried out to summarize their differences in terms of accuracy and efficiency when the three EBRBSs are used to address multi-class and multi-attribute classification problems. A useful guidance is obtained to distinguish the inherent features of the three EBRBSs.

(2) The procedure of applying the OVO decomposition strategy with five aggregation methods to enhance the three EBRBSs is showed for the first time. The corresponding investigation is further studied to analyze the influence of the OVO decomposition strategy on the three EBRBSs and their classification performance on multi-class problems.

(3) The procedure of applying square product, product, minimum, geometric mean, and sine overlap functions to enhance the three EBRBSs is showed for the first time. Meanwhile, another investigation is further studied to detail the influence of the five overlap functions on the three EBRBSs and their classification performance on multi-attribute problems.

The remainder of this paper is organized as follows: Section 2 reviews the related work on classifiers with decomposition strategy and overlap function. Section 3 provides an introduction of O-EBRBS, DRA-EBRBS, and Micro-EBRBS. Section 4 develops an integration framework to illustrate how decomposition strategy and overlap function enhance an EBRBS. Section 5 carries out an experimental study to analyze the performance on the enhanced EBRBSs. Finally, Section 6 concludes this work.

2. Related Works: Classifiers with Decomposition Strategy and Overlap Function

Throughout the past development of classical classifiers, decomposition strategies and overlap functions are useful to improve the classification performance of classifiers in terms of multi-class and multi-attribute problems. Many success attempts have been made to enhance conventional classifiers in the past decades for classification problems. These attempts can be divided into the following two aspects:

For the previous studies of classifiers with decomposition strategies, Fernández *et al*. [29] applied a pairwise learning methodology to construct a linguistic fuzzy multi-classifier systems, where the main idea of the pairwise learning methodology was based on the OVO strategy. They claimed that the proposed systems can obtain a better decision boundary in multi-class problems. Galar *et al*. [7] provided a survey of two commonly used binary decomposition strategies, including the OVO and one-versus-all (OVA) strategies, and also paid special attention to the combination of the outputs of binary classifiers. They suggested that the OVO strategy was able to improve the performance of some classifiers, such as support vector machine (SVM), decision tree (DT), Ripper, and *k*-nearest neighbor (KNN). Elkano *et al*. [30] combined the fuzzy association rule-based classification model for high-dimensional problems (FARC-HD) with the OVO and OVA strategies to improve classifier performance in multi-class problems. They concluded that the proposed model obtained competitive results in the comparison with three state-of-the-art fuzzy classifiers. Garcia *et al*. [31] utilized the OVO strategy to improve the performance of class noise filters so that the data quality in classification tasks could be enhanced because of detecting and removing error or noise data. Liu *et al*. [8] applied SVM with the OVO strategy for multi-class sentiment classification problems. The results showed the OVO strategy improved the performance of SVM better than other sentiment classification methods. Zhang *et al*., [32] empowered OVO decomposition strategy with ensemble learning for imbalanced multi-class datasets by using UnderBaging, SMOTEBagging, RUSBoost, SMOTEBoost, SMOTE+AdaBoost, and EasyEnsemble. The results indicated that the OVO strategy could significantly increase the effectiveness of ensemble learning for imbalanced multi-class problems. Recently, Liu and Jia [33] combined the OVO strategy and the instance-based learning for multi-class classification problems. The comparative studies obtained from seventeen benchmark multi-class datasets demonstrated that the proposed method had a better performance to predict the class of unseen instances against six well-established multi-class approaches.

For the previous studies of classifiers with overlap functions, Elkano *et al*. [30] addressed the problem caused by the low-confidence value involved in fuzzy reasoning methods, where the overlap functions were used in fuzzy rule-based classification system (FRBCS) to obtain more reasonable outputs of binary classifiers. Gomez *et al*. [13] justified the axiomatization used in the definition of overlap functions. The experimental results showed that overlap functions were able to obtain a better result than the commonly used product t-norm. Meanwhile, Elkano *et al*. [31] carried out an exhaustive study to investigate the influence of overlap functions on the FRBCS proposed by Chi *et al*. (Chi-FRBCS), structural learning algorithm in a vague environment (SLAVE), fuzzy unordered rule induction algorithm (FURIA), and FARC-HD. They concluded that the performance of overlap functions strongly depended on the learning process and rule structure of each FRBCS. De Miguel *et al*. [14] introduced the concept of a general overlap function and applied the general overlap function to define a new matching degree in FRBCS. They suggested that the general overlap function could improve the performance of a FRBCS in classification problems. Recently, Asmus *et al*. [34] studied the concept of n-dimensional interval-valued overlap functions with their representability and the definition of general interval-valued overlap functions with characterization.

The above literature reviews clearly show that decomposition strategies and overlap functions have positive influences on many conventional classifiers, even so the classifiers have the inherent capable of addressing multi-class and multi-attribute problems, in which FRBCS is one of commonly used classifiers in the existing studies to validate the improvements and influences made by decomposition strategies and overlap functions. Owing to the vigorous development in fuzzy fields, *e.g.*, fuzzy fractional derivative [35]-[38], FRBCS has been regarded as an outstanding representative rule-based system with superior performance over other machine learning methods [39]-[42]. Considering that EBRBS is a new developing rule-based system, it is worth studying the influence of decomposition strategies and overlap functions on the performance of EBRBSs. Hence, the present work aims to integrate the commonly used decomposition strategies and overlap functions into EBRBS and then study their influences on EBRBSs for multi-class and multi-attribute classification problems.

3. Three Representative EBRBSs for Classification Problems

In this section, the preliminary knowledge is introduced in Section 3.1 to show the components and differences of three kinds of representative EBRBSs, namely O-EBRBS, DRA-EBRBS, and Micro-EBRBS. Next, a brief description of the three EBRBSs is considered in Sections 3.2 to 3.4, respectively.

3.1. Preliminaries about Generic EBRBS

EBRBS is an advanced rule-based system extended from the belief rule-based system (BRBS) [44] by embedding belief structures into the rule antecedent of the belief rule that has already contained belief structures in rule consequent. Owing to belief structures, the rule in EBRB, which is so called extended belief rule, has a powerful representation scheme to simultaneously express fuzzy, random, and incomplete uncertainties caused by input and output data. Taking user knowledge modeling problem [45] for an example, the knowledge levels of users are categorized into *Very Low*, *Low*, *Middle*, and *High*, and one kind of belief structures for a user can be wrote as {(*Very Low*, 0.3), (*Low*, 0.6), (*Middle*, 0), (*High*, 0)}. Specifically, the knowledge level of the user is assessed to be *Very Low* with belief degree 0.3 or 30% possibility, *Low* with belief degree 0.6 or 60% possibility, and both *Middle* and *High* with belief degree 0 or 0% possibility. More importantly, due to 0.3 + 0.6 = 0.9 < 1.0, the remaining belief degree 1.0 - 0.9 = 0.1 or 10% possibility should be regarded as incomplete information to express uncertainty.

Based on the belief structure, when an EBRB is assumed to have M antecedent attributes and one consequent attribute, in which each antecedent attribute U_i (i=1,…, M) has J_i reference values $A_{i,j}$ (j=1,…, J_i) and consequent attribute D has N classes D_n (n=1,…, N), an extended belie rule in the EBRB can be written as:

$$\begin{aligned} R_k &: IF\ U_1\ is\ \{(A_{1,j},\alpha_{1,j}^k);\ j=1,\dots,J_1\}\wedge\cdots\wedge U_M\ \ is\ \{(A_{M,j},\alpha_{M,j}^k);\ j=1,\dots,J_M\}, \\ & THEN\ D\ is\ \{(D_n,\beta_n^k);n=1,\dots,N\},with\ \theta_k\ \ and\ \{\delta_1,\dots,\delta_M\} \end{aligned} \tag{1}$$

where $\alpha_{i,j}^k$ ($0\le\alpha_{i,j}^k\le1$) and β_n^k ($0\le\beta_n^k\le1$) denotes the belief degree in antecedent attribute U_i consequent attribute D; θ_k ($0\le\theta_k\le1$) is the weight of the kth (k=1,…, L) rule R_k, which represents the importance of R_k over other rules; δ_i ($0\le\delta_i\le1$) is the weight of antecedent attribute U_i, which represents the importance of U_i over other attributes.

On the basis of EBRB, an EBRBS consists of two main components: 1) EBRB generation scheme, which is a mechanism used to generate an EBRB based on expert knowledge and historical data; 2) EBRB Inference scheme, which is a mechanism used to classify input data using the extended belief rules stored in the EBRB. The basic framework of EBRBS is shown in Fig. 1.

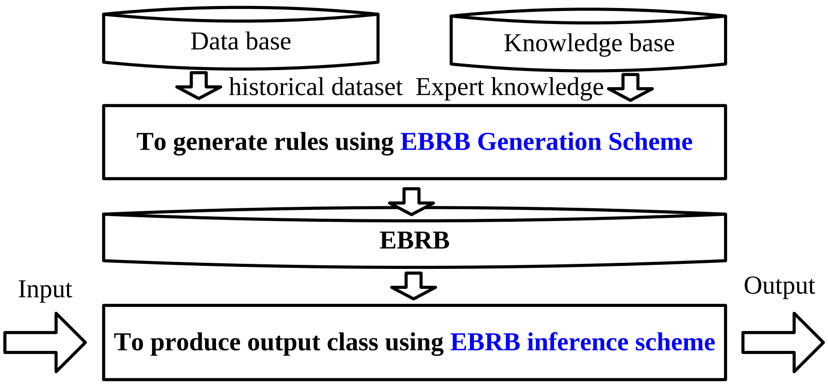


Fig. 1. Basic framework of EBRBS

From Fig. 1, the characteristics of EBRBS can be summarized as follows:

(1) A rule in EBRBS has a generic information representation, because it can express probabilistic, incomplete, and fuzzy uncertainties in both antecedent and consequent attributes.

(2) EBRBS is knowledge-driven, data-driven, or their combination system, because it can be automatically generated from data base and knowledge base, without requiring any time-consuming optimization process.

Additionally, since O-EBRBS, DRA-EBRBS, and Micro-EBRBS are considered in this study, the differences among these three EBRBSs in terms of EBRB generation scheme and EBRB inference scheme are illustrated in Fig. 2. The detailed steps of these components can be found in Section 3.2 to Section 3.4.

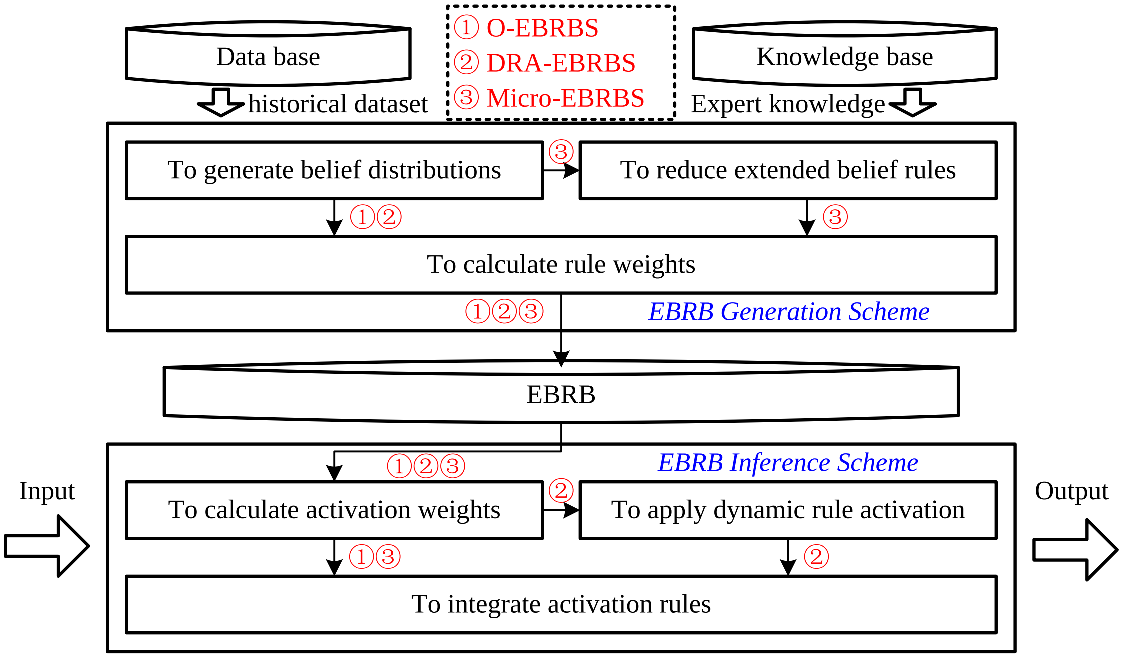


Fig. 2. Differences of three representative EBRBSs

Here, it is worth noting that data base and knowledge base are two prerequisites for constructing an EBRBS and they represent the collection of historical data and expert knowledge from a particular domain or certain problem, respectively. For example, in the case of user knowledge modeling problem, data base is constructed by using users' data collected by the user modeling server [45] and knowledge base is constructed by using the knowledge levels of users provided by domain professors, *e.g.*, the use of *Very Low*, *Low*, *Middle*, and *High* to describe the knowledge levels of users. Furthermore, all these expert knowledge can form the basic representation scheme of extended belief rules, *e.g.*, the use of the belief distribution {(*Very Low*, 0.3), (*Low*, 0.6), (*Middle*, 0), (*High*, 0)} to specify {(D_n, β_n^k); n=1,…, N} in Eq. (1) for constructing an EBRBS.

3.2. Introduction of O-EBRBS

O-EBRBS is the first EBRBS proposed in [18], which can be regarded as an extension of FRBCS and also an improvement of BRBS in terms of the flexibility of knowledge representation, *i.e.*, none of belief structures is used in fuzzy rules and belief structures are only used in the rule consequent of belief rules. Owing to this advantage, O-EBRBS has done a lot of effective and fruitful works on the domain of applications related with uncertainty information processing and modeling [23][25][46][47].

In order to construct an EBRB, the EBRB generation scheme of O-EBRBS should be performed based on the following two steps when the EBRB is assumed to have M antecedent attributes U_i (i=1,…, M) with J_i reference values $A_{i,j}$ (j=1,…, J_i) and one consequent attribute D with N classes D_n (n=1,…, N).

***Step 1*:** To generate belief distributions. Suppose $x_{k,i}$ is the kth (k=1,…, L) historical input data of the ith antecedent attribute U_i. A corresponding belief distribution of U_i can be generated using the utility-based equivalence transformation technique [44].

$$S(x_{k,i}) = \{(A_{i,j}, \alpha_{i,j}^k); j = 1,..., J_i\} \tag{2}$$

where

$$\alpha_{i,j}^{k}=\frac{u(A_{i,j+1})-x_{k,i}}{u(A_{i,j+1})-u(A_{i,j})} \text{ and } \alpha_{i,j+1}^{k}=1-\alpha_{i,j}^{k}, \text{ if } u(A_{i,j})\leq x_{k,i}\leq u(A_{i,j+1}) \tag{3}$$

$$\alpha_{i,t}^{k}=0,\ for\ t=1,...,J_i\ and\ t\neq j,j+1 \tag{4}$$

where $u(A_{i,j})$ denotes the utility value used for reference value $A_{i,j}$ in the ith antecedent attribute U_i; $\alpha_{i,j}^{k}$ denotes the belief degree of reference value $A_{i,j}$ in the kth rule generated from input data $x_{k,i}$.

Afterwards, when the kth historical output data y_k is assumed to be the jth class D_j (j=1,…, N), the belief distribution of consequent attribute D is expressed as follows:

$$S(y_k)=\{(D_n,\beta_n^k);n=1,...,N\} \tag{5}$$

where

$$\beta_n^k=\begin{cases}1; If\ n=j\\ 0; Otherwise\end{cases} \tag{6}$$

Step 2: To calculate rule weights. After transforming L historical input-output data pairs into L groups of belief distributions for M antecedent attributes and one consequent attribute, in which these belief distributions can form L extended belief rules, the similarity of rule antecedent (SRA) and the similarity of rule consequent (SRC) for these L extended belief rule can be calculated based on the following definitions:

Definition 1 (Distance of two belief distributions): Suppose there are two belief distributions $S(R_k, U_i)$={ $\alpha_{i,j}^{k}$; j=1,…, J_i} and $S(R_l, U_i)$={ $\alpha_{i,j}^{l}$; j=1,…, J_i} obtained from attribute U_i in rule R_k and R_l, thus the distance of $S(R_k, U_i)$ and $S(R_l, U_i)$ is as follows:

$$d(R_k,R_l,U_i)=\sqrt{\sum_{j=1}^{J_i}(\alpha_{i,j}^{k}-\alpha_{i,j}^{l})^2} \tag{7}$$

Definition 2 (Similarity of two belief distributions): On the basis of Definition 1, the similarity of two belief distributions $S(R_k, U_i)$ and $S(R_l, U_i)$ is as follows:

$$Sim(R_k,R_l,U_i)=1-\min\{1,d(R_k,R_l,U_i)\} \tag{8}$$

Thus, based on Definition 1 and Definition 2, the SRA and SRC of R_k and R_l (k, l=1,…, L) can be calculated by:

$$SRA(R_l,R_k)=\min_{i=1,...,M}\{Sim(R_k,R_l,U_i)\} \tag{9}$$

$$SRC(R_l,R_k)=Sim(R_k,R_l,D) \tag{10}$$

Then, the inconsistency degree of the kth extended belief rule is calculated as follows:

$$Incons(R_k)=\sum_{l=1,l\neq k}^{L}\left(1-\exp\left\{-\frac{\left(\frac{SRA(R_l,R_k)}{SRC(R_l,R_k)}-1\right)^2}{\left(\frac{1}{SRA(R_l,R_k)}\right)^2}\right\}\right) \tag{11}$$

Finally, the rule weight of the kth extended belief rule is calculated as follows:

$$\theta_k=1-\frac{Incons(R_k)}{\sum_{j=1}^{L}Incons(R_j)} \tag{12}$$

In order to classify any given input data, the EBRB inference scheme of O-EBRBS should be done using the following steps:

Step 1: To calculate activation weights. Suppose an input data $\boldsymbol{x}$=(x_1,…, x_M) is provided for O-EBRBS, each input x_i (i=1,…,

M) needs to be transformed into a belief distribution using Eqs. (3) to (4).

$$S(x_i) = \{(A_{i,j}, \alpha_{i,j}); j = 1,..., J_i\} \tag{13}$$

Next, the individual matching degree between the kth rule R_k and input data $\boldsymbol{x}$ regarding the ith antecedent attribute U_i, denoted as $S^k(x_i, U_i)$, is calculated based on Definition 1 as follows:

$$S^k(x_i, U_i) = \begin{cases} 1 - d(R_k, x_i, U_i), if\ d(R_k, x_i, U_i) < 1 \\ 0.5 \quad , else\ if\ U_i\ is\ nominal\ attribute \\ 0 \quad , otherwise \end{cases} \tag{14}$$

Finally, the activation weight of the kth extended belief rule, denoted as w_k, is calculated by

$$w_k = \frac{\theta_k \prod_{i=1}^{M} (S^k(x_i, U_i))^{\bar{\delta}_i}}{\sum_{l=1}^{L} \theta_l \prod_{i=1}^{M} (S^l(x_i, U_i))^{\bar{\delta}_i}}, \bar{\delta}_i = \frac{\delta_i}{\max_{t=1,...,M}\{\delta_t\}} \tag{15}$$

where θ_k is the weight of the kth rule; δ_i is the weight of the ith antecedent attribute; Here, it worth noting that if w_k is greater than 0, the kth rule should be regarded as an activated rule to produce an output class.

***Step 2*:** To integrate activated rules. Suppose that L extended belief rules are all activated for the input data $\boldsymbol{x}$. Hence, all these L activated rules should be integrated using the following analytical ER algorithm [48]:

$$\beta_n = \frac{\prod_{k=1}^{L} (w_k \beta_n^k + 1 - w_k \sum_{i=1}^{N} \beta_i^k) - \prod_{k=1}^{L} (1 - w_k \sum_{i=1}^{N} \beta_i^k)}{\sum_{i=1}^{N} \prod_{k=1}^{L} (w_k \beta_i^k + 1 - w_k \sum_{j=1}^{N} \beta_j^k) - (N-1) \prod_{k=1}^{L} (1 - w_k \sum_{j=1}^{N} \beta_j^k) - \prod_{k=1}^{L} (1 - w_k)} \tag{16}$$

where β_n is the integrated belief degree, which represents the possibility of the input data $\boldsymbol{x}$ to be the nth class D_n.

Next, the output class of O-EBRBS is obtained by seeking the one with the greatest belief degree.

$$f(\boldsymbol{x}) = D_n, n = \arg\max_{i=1,...,N}\{\beta_i\} \tag{17}$$

3.3. Introduction of DRA-EBRBS

DRA-EBRBS is an improved EBRBS in term of activation rules determination [15]. As shown in Fig. 2, the main difference between O-EBRBS and DRA-EBRBS is that the dynamic rule activation is considered to determine activation rules in the EBRB inference scheme. The detailed steps of the dynamic rule activation are introduced as follows:

Firstly, by introducing a parameter λ and according to Eq. (14), new individual matching degree is calculated by

$$S_\lambda^k(x_i, U_i) = (S^k(x_i, U_i))^\lambda \tag{18}$$

where the parameter λ may affect the individual matching degree in two different ways: 1) higher λ values penalize activated rules with low activation weights and even zero activation weight; 2) lower λ values reward activated rules with high activation weights. More specifically, when λ =0, the activation weights is equal to rule weights because of $S_\lambda^k(x_i, U_i)$ =1.

Secondly, based on the new individual matching degree shown in Eq. (18), new activation weight of rule R_k, denoted as w_k, is calculated as well. When w_k is greater than 0, rule R_k should be put into the set Δ_λ, namely $\Delta_\lambda = \Delta_\lambda \cup R_k$.

Thirdly, a function, denoted as $C(\Delta_\lambda)$, is utilized to measure the performance of λ in term of consistency in the set Δ_λ, in which the function $C(\Delta_\lambda)$ is defined as follows:

$$C(\Delta_\lambda) = \frac{\max_{n=1,...,N}\{C_n\}}{|\Delta_\lambda|} \tag{19}$$

where $C(\Delta_\lambda)$ represents the maximum number of rules with the maximum belief degree in the same class divided by the total number of rules; C_n represents the number of rules with the maximum belief degree in the nth class D_n and its value is given by:

$$C_n = | D_n ; n = \arg\max_{i=1,\ldots,N} \{\beta_i^k\}; R_k \in \Delta_\lambda | \tag{20}$$

Finally, by searching for the range of λ to obtain the maximum value of $C(\Delta_\lambda)$, the relevant set Δ_λ is regarded as the final set of activation rules and all these activation rules are used to generate final output class using Eqs. (16) and (17).

3.4. Introduction of Micro-EBRBS

Micro-EBRBS is another improved EBRBS in term of ineffective rules reduction [11]. As observed in Fig. 2, the difference between O-EBRBS and Micro-EBRBS is that rule reduction is considered to downsize EBRB in the EBRB generation scheme. The detailed steps of rule reduction are introduced as follows:

Firstly, suppose an EBRB has L extended belief rules, M antecedent attributes U_i (i=1,…, M) with J_i reference values $A_{i,j}$ (j=1,…,J_i), and one consequent attribute D with N classes D_n (n= 1,…, N). Hence, $\prod_{i=1}^{M} J_i$ division domains { $D(A_{1,j_1},\ldots,A_{M,j_M})$; j_i=1,…, J_i; i=1,…, M} can be generated by dividing global input space into multiple local input spaces based on the combination of all reference values for each antecedent attribute.

Secondly, all extended belief rules are mapped into division domains according to the following map function:

$$R_k \to D(A_{1,j_1},\ldots,A_{M,j_M}); j_i = \arg\max_{j=1,\ldots,J_i} \{\alpha_{i,j}^k\}; k=1,\ldots,L; i=1,\ldots,M \tag{21}$$

where the map function in Eq. (21) means the collection of the rules with the maximum belief degree in the same reference values.

Thirdly, for the division domain which has one rule at least, all rules in the same division domain are used to generate a new extended belief rule, in which the belief degrees of antecedent and consequent attributes in the new rule are calculated as follows:

$$\overline{\alpha}_{i,j}^{l} = \frac{\sum_{k=1}^{L_l} \alpha_{i,j}^k}{L_l}, \overline{\beta}_n^{l} = \frac{\sum_{k=1}^{L_l} \beta_n^k}{L}; i=1,\ldots,M; j=1,\ldots,J_i; l=1,\ldots,\prod_{i=1}^{M} J_i; n=1,\ldots,N \tag{22}$$

where L_l denotes the number of rules in the lth division domain; $\alpha_{i,j}^k$ and β_n^k denote the belief degrees used in the kth extended belief rule; $\overline{\alpha}_{i,j}^{l}$ denotes the belief degree of the jth reference value of the ith antecedent attribute at the lth new extended belief rule. $\overline{\beta}_n^{l}$ denotes the belief degree of the nth class at the lth new extended belief rule.

Finally, the rule weight of all new extended belief rules is calculated using Eqs. (9) to (12) to construct a downsized EBRB for Micro-EBRBS. Note that this makes sense that the EBRB in Micro-EBRB is a micro version comparing to O-EBRBS.

4. Enhancing EBRBSs based on Decomposition Strategy and Overlap Function

In this section, the procedures for enhancing EBRBSs are developed to illustrate how decomposition strategies and overlap functions work, in which the basic framework of the enhanced EBRBSs is shown in Section 4.1 and its two procedures, namely modeling binary EBRBS and modeling conjunctive relationship, are detailed in Sections 4.2 and 4.3, respectively.

4.1. Basic Framework of the EBRBS with Decomposition Strategy and Overlap Function

To clearly illustrate how decomposition strategies and overlap functions work, an enhanced EBRBS is proposed on the basis of using a decomposition strategy and an overlap function to enhance the original EBRB generation scheme and EBRB inference scheme, where Fig. 3 shows the main framework of the enhanced EBRBS.

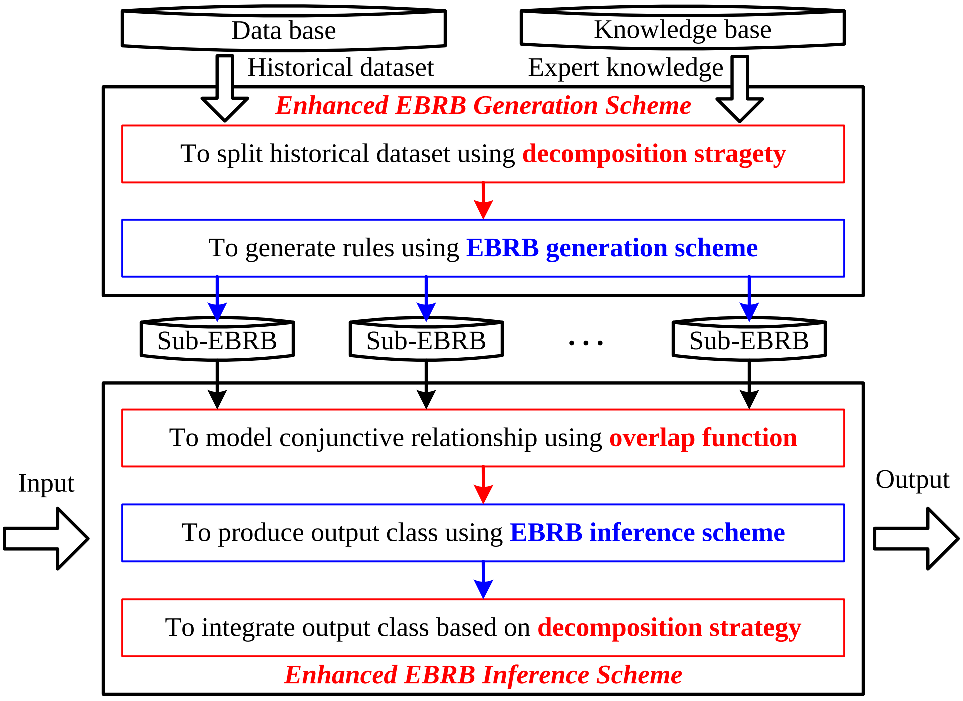


Fig. 3. Basic framework of enhanced EBRBSs

According to Fig. 3, the basic framework of enhanced EBRBSs can be described as follows:

(1) Enhanced EBRB generation scheme. This is an enhanced scheme to generate EBRB based on knowledge base and data base. Comparing to the original EBRB generation scheme, historical dataset should be divided into multiple sub-datasets and all these sub-datasets are further utilized to generate multiple sub-EBRBs. Hence, this scheme mainly includes two steps: 1) to split historical dataset using a decomposition strategy; 2) to generate rules using the original EBRB generation scheme. Noting that the details of the first step can be found in Section 4.2 and the second step is shown in Section 3.

(2) Enhanced EBRB inference scheme. This is another enhanced scheme used to produce an output class based on multiple sub-EBRBs. Comparing to the original EBRB inference scheme, overlap functions are used to model the conjunctive relationship among antecedent attributes in each sub-EBRB and the output class derived from each sub-EBRB should be integrated to produce a final output class. Thus, this scheme mainly includes three steps: 1) to model conjunctive relationship using an overlap function; 2) to produce an output class using the original EBRB inference scheme; 3) to integrate the output class based on decomposition strategy. Noting that the details of the first step can be found in Section 4.3, the second step is shown in Section 3 and the details of the third step are in Section 4.2.

Here, it is worth noting that the use of decomposition strategies and overlap functions can improve the classification accuracy of EBRBSs, but they have to come with the price of additional computing complexity, *i.e.*, the OVO decomposition strategy can divide one complex problem into multiple simple sub-problems, but this also means that the EBRB generation scheme and EBRB inference scheme should be performed for multiple times when handling these simple sub-problems.

4.2. Modeling Binary EBRBSs Using OVO Strategy

The kernel of decomposition strategies is to divide a complex multi-class classification problem into multiple simper ones so that it can be independently handled by multiple classifiers. Hence, decomposition strategies are not only useful to enhance the performance of the classifiers which are just able to address simple classification problems, but also have an inherent multi-class support. Based upon this viewpoint, the OVO strategy is used to construct binary EBRBSs for multi-class classification problems. The illustration is shown in Fig. 4.

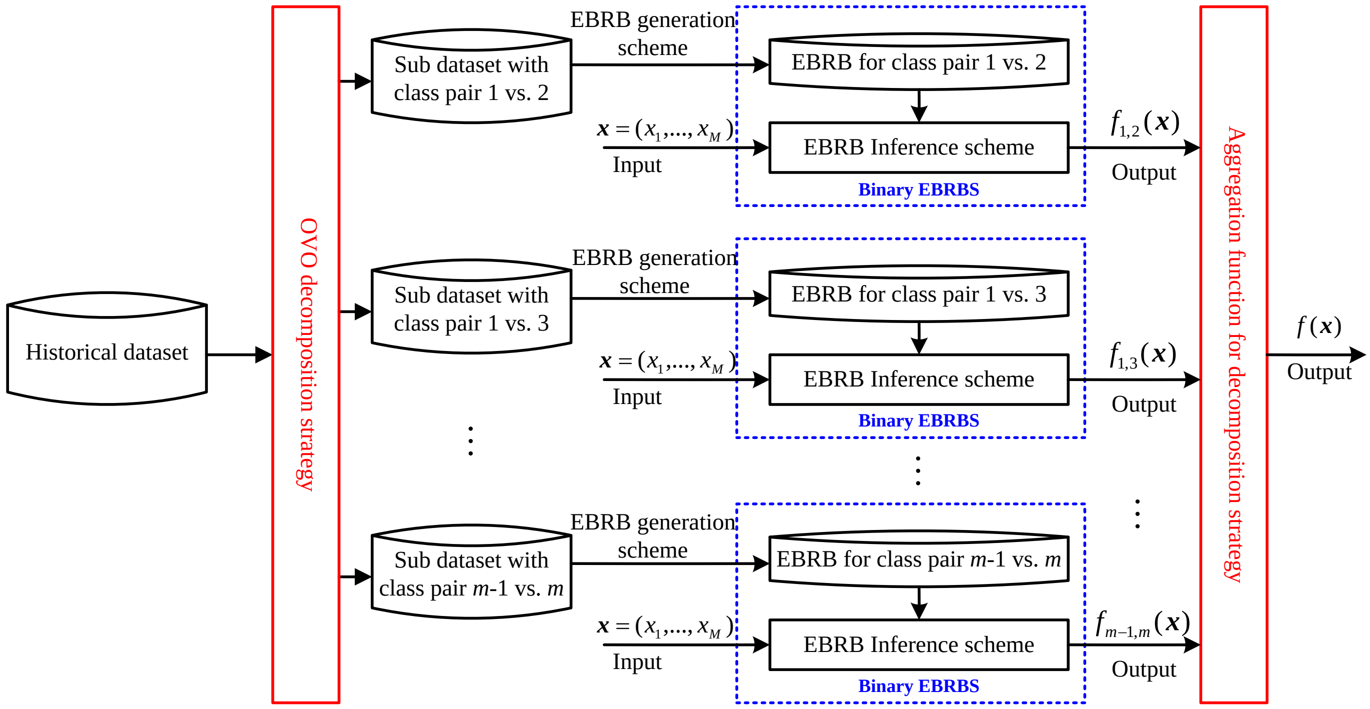


Fig. 4. Illustration of EBRBS with decomposition strategy

It is clear from Fig. 4 that the OVO strategy divides a historical dataset (with N classes) into N * (N - 1) / 2 binary sub-datasets (with two classes). Each binary sub-dataset should be used to construct an EBRB according to the EBRB generation scheme shown in Section 3, namely binary EBRBS. For convenient discussion, the integrated belief degrees produced by the binary EBRBS, which is responsible for the classes D_i and D_j, are denoted by $\beta_{i,j}\,(0 \le \beta_{i,j} \le 1)$ and $\beta_{j,i}\,(0 \le \beta_{j,i} \le 1)$.

As a result, for a given input data $\boldsymbol{x}$, the integrated belief degrees obtained from all binary EBRBSs can be as follows:

$$\boldsymbol{\beta}(\boldsymbol{x}) = \begin{pmatrix} - & \beta_{1,2} & \cdots & \beta_{1,N} \\ \beta_{2,1} & - & \cdots & \beta_{2,N} \\ \vdots & \vdots & & \vdots \\ \beta_{N,1} & \beta_{N,2} & \cdots & - \end{pmatrix} \tag{23}$$

where $\beta_{i,j}$ and $\beta_{j,i}$ (i, j=1,…, N) are the outputs of the binary EBRBS constructed by using the sub-dataset which only includes two classes D_i and D_j and therefore they represent the belief degree of the ith class and the jth class, respectively.

Since each binary EBRBS is independent for classifying input data $\boldsymbol{x}$, the integrated belief degrees shown in Eq. (23) needs to be normalized in order to have all confidence degrees within the same range of values. The normalization of the integrated belief degrees $\hat{\boldsymbol{\beta}}(\boldsymbol{x}) = \{\hat{\beta}_{i,j}; i, j = 1,..., N\}$ is performed as follows:

$$\hat{\beta}_{i,j} = \begin{cases} 0.5, & if\ \beta_{i,j} = \beta_{j,i} = 0 \\ \dfrac{\beta_{i,j}}{\beta_{i,j} + \beta_{j,i}}, & otherwise \end{cases} \tag{24}$$

Therefore, the final output class of all binary EBRBSs can be obtained using the following aggregation functions:

(1) Voting strategy (VS) function [49]. Each binary EBRBS gives a vote for the predicted class by using value 1. The class having the maximum value is regarded as the final output class:

$$f(\boldsymbol{x}) = D_n, n = \arg\max_{i=1,...,N} \{\textstyle\sum_{j=1; j\neq i}^{N} s_{i,j}\} \tag{25}$$

where

$$s_{i,j}=\begin{cases}1; & If\ \hat{\beta}_{i,j}\geq\hat{\beta}_{j,i}\\ 0; & Otherwise\end{cases} \tag{26}$$

(2) Weighted Voting (WV) function [50]. Each binary EBRBS gives a vote for each class by using belief degrees. The class having the maximum belief degree is regarded as the final output class:

$$f(\boldsymbol{x})=D_n, n=\arg\max_{i=1,...,N}\{\textstyle\sum_{j=1;j\neq i}^{N}\hat{\beta}_{i,j}\} \tag{27}$$

(3) WinWV (WWV) function [43]. Each binary EBRBS gives a vote for the predicted class by using belief degrees. The class having the maximum belief degree is regarded as the final output class:

$$f(\boldsymbol{x})=D_n, n=\arg\max_{i=1,...,N}\{\textstyle\sum_{j=1;j\neq i}^{N}s_{i,j}\} \tag{28}$$

where

$$s_{i,j}=\begin{cases}\hat{\beta}_{i,j}; & If\ \hat{\beta}_{i,j}\geq\hat{\beta}_{j,i}\\ 0; & Otherwise\end{cases} \tag{29}$$

(4) Non-dominance criteria (NC) function [51]. Each binary EBRBS gives a vote for the predicted class by using non-dominance degree. The class having the maximum degree is regarded as the final output class:

$$f(\boldsymbol{x})=D_n, n=\arg\max_{i=1,...,N}\{1-\max_{j=1,...,N}\{t_{j,i}\}\} \tag{30}$$

where

$$t_{i,j}=\begin{cases}\hat{\beta}_{i,j}-\hat{\beta}_{j,i}; & If\ \hat{\beta}_{i,j}\geq\hat{\beta}_{j,i}\\ 0; & Otherwise\end{cases} \tag{31}$$

(5) Learning valued preference for classification (LC) function [29]. Each binary EBRBS gives a vote for each class by using strict preference, conflict degree, and ignorance degree. The class having the maximum value is regarded as the final output class:

$$f(\boldsymbol{x})=D_n, n=\arg\max_{i=1,...,N}\left\{\sum_{j=1;j\neq i}^{N}\left(P_{i,j}+\frac{C_{i,j}}{2}+\frac{N_i}{N_i+N_j}I_{i,j}\right)\right\} \tag{32}$$

where N_i is the number of sample data related to the class C_i and this parameter is used as the unbiased estimate of the class probability; $C_{i,j}$ denotes the conflict degree, namely the degree to which the ith and the jth classes are supported, and $C_{i,j}=\min\{\hat{\beta}_{i,j}, \hat{\beta}_{j,i}\}$; $P_{i,j}$ denotes the strict preference for the ith and the jth classes, and $P_{i,j}=\hat{\beta}_{i,j}-C_{i,j}$, $I_{i,j}$ denotes ignorance degree, namely the degree to which none of the ith and the jth classes is supported, and $I_{i,j}=1-\max\{\hat{\beta}_{i,j}, \hat{\beta}_{j,i}\}$. It is worth noting that at least one of these two degrees is zero and $P_{i,j}+P_{j,i}+C_{i,j}+I_{i,j}=1$.

4.3. Modeling Conjunctive Relationship Using Overlap Function

The earliest overlap functions were used for image processing in an n-dimensional space to classify the pixels of object and background. Considering that these overlap functions satisfy the similar properties of the functions widely used to model the conjunction relationship in rule-based systems, several commonly used overlap functions are considered for EBRBS. For the sake of convenience, both $(S^k(x_i,U_i))^{\bar{\delta}_i}$ in Eq. (14) and $(S_\lambda^k(x_i,U_i))^{\bar{\delta}_i}$ in Eq. (18) are denoted by S_i^k ($0\leq S_i^k\leq 1$). As a result, the enhanced activation weight calculation using overlap functions can be illustrated as Fig. 5.

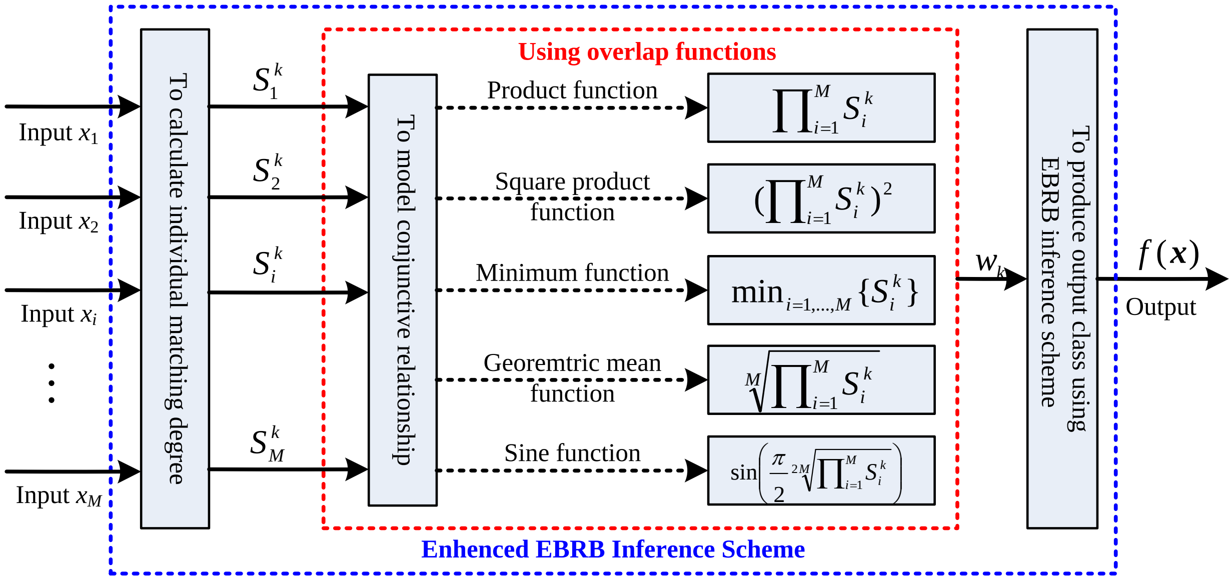


Fig. 5. Illustration of activation weight calculation with overlap functions

According to Fig. 5, the activation weight calculation with an overlap function can be expressed as follows:

(1) Product (PR) function: the produced value is based on the product of S_i^k, namely,

$$w_k = \frac{\theta_k O(S_1^k,\dots,S_M^k)}{\sum_{l=1}^{L}\theta_l O(S_1^l,\dots,S_M^l)} \tag{33}$$

where θ_k is the weight of the kth extended belief rule and its calculation is shown in Eq. (12); $O(S_1^k,\dots,S_M^k)$ is the function of aggregating S_i^k (i=1,…, M) and its detailed formula is shown as follows:

$$O(S_1^k,\dots,S_M^k) = \prod_{i=1}^{M} S_i^k \tag{34}$$

In addition, the following overlap functions are also applied to calculate the activation weights for EBRBS:

(2) Minimum (MI) function: the produced value is based on the minimum of S_i^k, namely,

$$O(S_1^k,\dots,S_M^k) = \min_{i=1,\dots,M}\{S_i^k\} \tag{35}$$

(3) Square product (SP) function: the produced value is based on the square product of S_i^k, namely,

$$O(S_1^k,\dots,S_M^k) = \left(\prod_{i=1}^{M} S_i^k\right)^2 \tag{36}$$

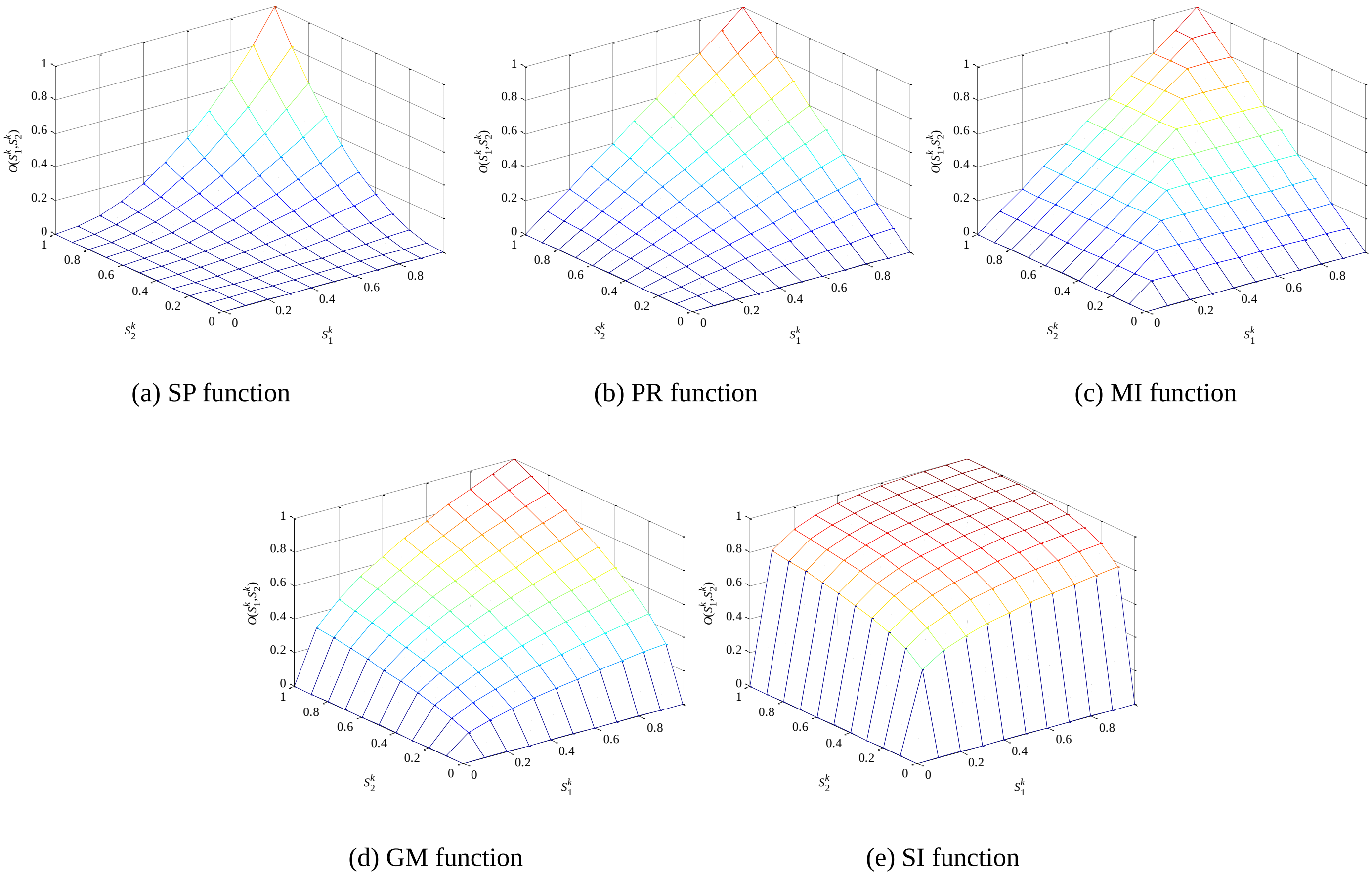
(4) Geometric mean (GM) function: the produced value is based on the geometric mean of S_i^k, namely,

$$O(S_1^k,\dots,S_M^k) = \sqrt[M]{\prod_{i=1}^{M} S_i^k} \tag{37}$$

(5) Sine (SI) function: this function is to produce the value higher than means, namely,

$$O(S_1^k,\dots,S_M^k) = \sin\left(\frac{\pi}{2}\sqrt[2M]{\prod_{i=1}^{M} S_i^k}\right) \tag{38}$$

In order to show the difference of the five overlap functions, suppose that there are two antecedent attributes (*i.e.*, M =2) and the range of S_1^k and S_2^k is within interval [0, 1]. The value of $O(S_1^k, S_2^k)$ produced by the five overlap functions in the 2-dimensional space are therefore shown in Fig. 6.



(a) SP function (b) PR function (c) MI function

(d) GM function (e) SI function

Fig. 6. Values produced by five kinds of 2-dimensional overlap functions

As we can observe from Fig. 6, three features are summarized as follows:

(1) The range of $O(S_1^k, S_2^k)$ regarding five overlap functions in 2-dimensional space is within interval [0, 1].

(2) The ranking order of five overlap functions to produce smaller value is SP $\succ$ PR $\succ$ MI $\succ$ GM $\succ$ SI.

For example, when the value of S_1^k and S_2^k is 0.2 and 0.3, the values of $O(S_1^k, S_2^k)$ are therefore 0.0036, 0.06, 0.2, 0.245, and 0.777 for the SP, PR, MI, GM, and SI functions, respectively. It is evident that all these produced values coincide with the above two features, such as 0 < 0.0036 < 0.06 < 0.2 < 0.245 < 0.777 < 1.

5. Experimental Study

In this section, an experimental study is carried out according to the following aspects: Section 5.1 introduces datasets and experiment conditions; Section 5.2 provides a real scenario to test O-EBRBS, DRA-EBRBS, and Micro-EBRBS; Sections 5.2 compares the experimental results obtained from the EBRBSs; Sections 5.3 and 5.4 analyze the influence of OVO decomposition strategy and overlap functions on EBRBSs; Section 5.5 discusses the reasons of influence for EBRBSs.

5.1. Datasets and Experiment Conditions

This subsection aims at introducing the classification datasets and experiment conditions used in the experimental study of investigating the influence of decomposition strategies and overlap functions on EBRBSs. Firstly, twenty datasets obtained from the KEEL dataset repository [27] and their detailed descriptions are shown in Table 1, in which these descriptions mainly include number of data (#Data), number of attributes (#Attrs) , number of nominal attributes (#Nomi), and number of classes (#Classes).

Table 1. Statistics on twenty classification datasets

No.	Datasets	#Data	#Attrs	#Nomi	#Classes	No.	Datasets	#Data	#Attrs	#Nomi	#Classes
1	Autos	159	25	10	6	11	Penbased	1100	16	0	10
2	Car	1728	6	0	4	12	Seeds	210	7	0	3
3	Contraceptive	1473	9	6	3	13	Segment	2310	19	60	7
4	Ecoli	336	7	0	8	14	Tae	151	5	2	3
5	Flare	1066	11	3	6	15	Thyroid	720	21	0	3
6	Glass	214	9	0	7	16	Vehicle	846	18	0	4
7	Iris	150	4	11	3	17	Vowel	990	13	0	11
8	Knowledge	403	5	0	4	18	Wine	178	13	0	3
9	Lymphography	148	18	15	4	19	Yeast	1484	8	0	11
10	Nursery	1296	8	8	5	20	Zoo	101	16	16	7

In order to obtain the expected performance of enhanced EBRBSs over an entire dataset and correct dataset shift, the 5-fold distribution optimally balanced cross validation [52] is considered in the experimental study, where each dataset is divided into 5 blocks, with 4 blocks as training data, namely historical input-output data pairs, and the remaining block as testing data, namely the input data needed to be classified using a classifier.

The most visualized metrics, including average accuracy and number of rules and activated rules, are used to quantify the expected performance of enhanced EBRBSs. Considering sometime it is not easy to extract justified conclusions based on those visualized metrics, the Aligned Friedman and Holm tests [28] are applied to provide statistical supports in the experimental study. Specifically, $\alpha = 0.1$ is used as the level of significance in all cases to show statistical differences.

5.2. Real scenario to test O-EBRBS, DRA-EBRBS, and Micro-EBRBS

This subsection aims to test O-EBRBS, DRA-EBRBS, and Micro-EBRBS by using a real classification dataset collected from user knowledge modeling problem [45], which is also called Knowledge shown in Table 1 and its subject is about the students' knowledge status and the subject of Electrical DC Machines. More specifically, the dataset of user knowledge modeling problem include four antecedent attributes, namely the degree of study time for goal object materials (U_1), the degree of repetition number of user for goal object materials (U_2), the degree of study time of under for related objects with goal object (U_3), the exam performance of user for related objects with goal object (U_4), and the exam performance of user for goal objects, as well as one consequent attribute named the knowledge level of user (D), which has four levels, namely *Very Low* (D_1), *Low* (D_2), *Middle* (D_3), and *High* (D_4). On the basis of the above antecedent and consequent attributes, 403 data were collected from the undergraduate students of department of Electrical Education of Gazi University in the 2009 semester, and the value range of the four antecedent attributes is within interval [0, 0.99], [0, 0.9], [0, 0.95], [0, 0.99], and [0, 0.99], respectively.

(1) The use of the EBRB generation scheme to construct O-EBRBS, DRA-EBRBS, and Micro-EBRBS.

Firstly, the number of reference values and their utility values should be given according to expert knowledge. Hence, it is assumed that there are three reference values ($A_{i,j}$, i=1,…, 5; j=1,…, 3) for each attribute and their utility values $u(A_{i,j})$ are evenly distributed within the definition interval of each attribute, as shown in Table 2.

Table 2. Utility values of three reference values for five attributes in user knowledge modeling

$u(A_{i,j})$	U_1	U_2	U_3	U_4	U_5
$u(A_{i,1})$	0.000	0.000	0.000	0.000	0.000
$u(A_{i,2})$	0.495	0.450	0.475	0.495	0.495
$u(A_{i,3})$	0.990	0.900	0.950	0.990	0.990

(1.1) To generate belief distributions for O-EBRBS, DRA-EBRBS, and Micro-EBRBS.

Based on the utility values shown in Table 2, 322 training data can be used to generate belief distributions according to ***Step 1*** of the EBRB generation scheme shown in Section 3.2. For example, when one input-output data pair is <$x_{k,1}$=0.10, $x_{k,2}$=0.27, $x_{k,3}$=0.31, $x_{k,4}$=0.29, $x_{k,5}$=0.65, y_k=D_3>, the belief distribution regarding $x_{k,1}$=0.10 is calculated using Eqs. (3) and (4) as follows:

$$\alpha_{1,1}^{k}=\frac{u(A_{1,2})-x_{k,1}}{u(A_{1,2})-u(A_{1,1})}=\frac{0.495-0.1}{0.495-0}=0.798 \tag{39}$$

$$\alpha_{1,2}^{k}=1-\alpha_{1,1}^{k}=1-0.798=0.202 \tag{40}$$

$$\alpha_{1,3}^{k}=0 \tag{41}$$

From Eqs. (39) to (41), the belief distribution regarding $x_{k,1}$=0.10 is expressed as $S(x_{k,1})$={($A_{1,1}$, 0.798), ($A_{1,2}$, 0.202), ($A_{1,3}$, 0)}. Similarly, the belief distribution of $x_{k,2}$=0.27, $x_{k,3}$=0.31, $x_{k,4}$=0.29, $x_{k,5}$=0.65, and y_k=D_2 can be obtained using Eqs. (2) to (6) and they are $S(x_{k,2})$={($A_{2,1}$, 0.4), ($A_{2,2}$, 0.6), ($A_{2,3}$, 0)}, $S(x_{k,3})$={($A_{3,1}$, 0.347), ($A_{3,2}$, 0.653), ($A_{3,3}$, 0)}, $S(x_{k,4})$={($A_{4,1}$, 0.414), ($A_{4,2}$, 0.586), ($A_{4,3}$, 0)}, $S(x_{k,5})$={($A_{5,1}$, 0), ($A_{5,2}$, 0.687), ($A_{5,3}$, 0.313)}, and $S(y_k)$={(D_1, 0), (D_2, 0), (D_3, 1), (D_4, 0)}, respectively. As a consequence, a total of 322 belief distributions can be transformed from 322 training data for each attribute.

(1.2) To calculate rule weights for O-EBRBS and DRA-EBRBS.

According to ***Step 2*** of the EBRB generation scheme shown in Section 3.2, 322 belief distributions can form 322 extended belief rules and all these belief distributions are further used to calculate SRA and SRC for each rule by using Eqs. (9) and (10). Finally, the weight of 322 rules can be calculated using Eqs. (11) to (12) and they are shown in Fig. 7.

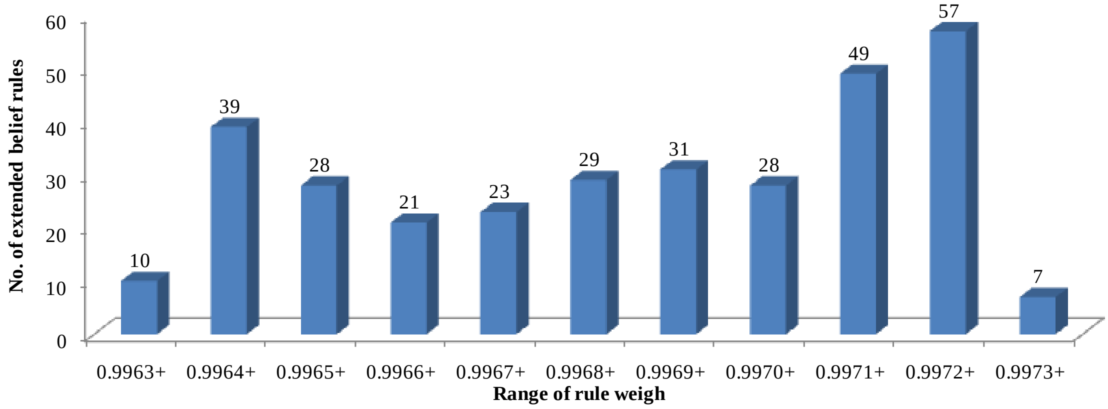


Fig. 7. Range of rule weight for 322 extended belief rules

From Fig. 7, it can be found that all rule weights are within interval [0.9963, 0.9974], which means that all extended belief rules are important rules for user knowledge modeling. More detailedly, the top two maximum numbers of extended belief rules are 57 and 49, and they are within interval [0.9971, 0.9972) and [0.9972, 0.9973), respectively; the two two minimum numbers of extended belief rules are 7 and 10, and they are within interval [0.9963, 0.9964) and [0.9973, 0.9974], respectively.

(1.3) To reduce extended belief rules for Micro-EBRBS.

Continuing with the dataset of user knowledge modeling to construct Micro-EBRBS, according to Eqs. (21) to (22) shown in Section 3.4, it needs to combine extended belief rules which fall in the same division domain. For example, there are two belief distributions {($A_{1,1}$, 0.798), ($A_{1,2}$, 0.202), ($A_{1,3}$, 0)} and {($A_{1,1}$, 0.636), ($A_{1,2}$, 0.364), ($A_{1,3}$, 0)} for two rules R_1 and R_2 which only have one antecedent attribute, it needs to therefore generate a new belief distribution {($A_{1,1}$, 0.722), ($A_{1,2}$, 0.278), ($A_{1,3}$, 0)} because of R_1, $R_2 \rightarrow D(A_{1,1})$. Based upon this viewpoint, parts of extended belief rules in O-EBRBS can be reduced to finally generate 110 new extended belief rules for Micro-EBRBS.

(2) The use of the EBRB inference scheme of O-EBRBS, DRA-EBRBS, and Micro-EBRBS to classify input data.

For the sake of convenience, suppose that O-EBRBS, DRA-EBRBS, and Micro-EBRBS have the same extended belief rules and they are shown in Table 3. One input data needed to be classified is $\boldsymbol{x}$=<x_1=0.19, x_2=0.38, x_3=0.38, x_4=0.49, x_5=0.45>.

Table 3. Three extended belief rules in O-EBRBS, DRA-EBRBS, and Micro-EBRBS

R_k	θ_k	U_1 (δ_1=1) ($\alpha^k_{1,1}, \alpha^k_{1,2}, \alpha^k_{1,3}$)	U_2(δ_2=1) ($\alpha^k_{2,1}, \alpha^k_{2,2}, \alpha^k_{2,3}$)	U_3(δ_3=1) ($\alpha^k_{3,1}, \alpha^k_{3,2}, \alpha^k_{3,3}$)	U_4(δ_4=1) ($\alpha^k_{4,1}, \alpha^k_{4,2}, \alpha^k_{4,3}$)	U_5(δ_5=1) ($\alpha^k_{5,1}, \alpha^k_{5,2}, \alpha^k_{5,3}$)	D ($\beta^k_1, \beta^k_2, \beta^k_3, \beta^k_4$)
R_1	0.987	(0.636, 0.364, 0)	(0.311, 0.689, 0)	(0.326, 0.674, 0)	(0.152, 0.848, 0)	(0.434, 0.566, 0)	(0, 0, 1, 0)
R_2	0.789	(0.782, 0.202, 0)	(0.400, 0.600, 0)	(0.347, 0.653, 0)	(0.414, 0.586, 0)	(0, 0.687, 0.313)	(0, 0, 1, 0)
R_3	0.234	(0.879, 0.121, 0)	(0.356, 0.644, 0)	(0.263, 0.737, 0)	(0, 0.465, 0.535)	(0.495, 0.505, 0)	(0, 1, 0, 0)

(2.1) To calculate activation weights and integrate activated rules for O-EBRBS and Micro-EBRBS.

According to ***Step 1*** of the EBRB inference scheme shown in Section 3.2, the belief distributions of $\boldsymbol{x}$ should be calculated and they are $S(x_1)$={($A_{1,1}$, 0.616), ($A_{1,2}$, 0.384), ($A_{1,3}$, 0)}, $S(x_2)$={($A_{2,1}$, 0.156), ($A_{2,2}$, 0.844), ($A_{2,3}$, 0)}, $S(x_3)$={($A_{3,1}$, 0.200), ($A_{3,2}$, 0.800), ($A_{3,3}$, 0)}, $S(x_4)$={($A_{4,1}$, 0.010), ($A_{4,2}$, 0.990), ($A_{4,3}$, 0)}, and $S(x_5)$={($A_{5,1}$, 0.091), ($A_{5,2}$, 0.909), ($A_{5,3}$, 0)}. Owing to the fact that O-EBRBS and Micro-EBRBS have the same processes of calculating activation weights, Table 4 provides the individual matching degrees of five antecedent attributes $S^k(x_i, U_i)$ (i=1,…, 5) and the activation weights of three extended belief rules w_k (k= 1,…, 3). From Table 4, it is clear that all three extended belief rules should be activated to classify input data $\boldsymbol{x}$, but R_1 is the most important activated rules comparing to R_2 and R_3.

Table 4. Individual matching degrees , and activation weights in O-EBRBS and Micro-EBRBS

	$S^k(x_1, U_1)$	$S^k(x_2, U_2)$	$S^k(x_3, U_3)$	$S^k(x_4, U_4)$	$S^k(x_5, U_5)$	w_k
R_1	0.972	0.780	0.822	0.799	0.515	0.737
R_2	0.754	0.654	0.792	0.429	0.606	0.233
R_3	0.628	0.717	0.911	0.250	0.429	0.030

According to ***Step 2*** of the EBRB inference scheme shown in Section 3.2, the belief distribution of consequent attribute in R_1, R_2 and R_3 should be integrated using the analytical ER algorithm. Finally, the integrated belief distribution is therefore {(D_1, 0), (D_2, 0.008), (D_3, 0.992), (D_4, 0)} and the output class is D_3 or *Middle*.

(2.2) To calculate activation weights and integrate activated rules for DRA-EBRBS.

Based on the dynamic rule activation method for DRA-EBRBS shown in Section 3.3, it needs to consider the parameter λ to calculate individual matching degrees and select consistent activated rules using Eqs. (18) to (20), *i.e.*, when λ=1, individual matching degrees $S^k_\lambda(x_i, U_i)$ and activation weights w_k of DRA-EBRBS are the same as the values shown in Table 4. As a result, the rule set is Δ_λ={R_1, R_2, R_3} and the function is $C(\Delta_\lambda)$=0.667; when λ=4, individual matching degrees $S^k_\lambda(x_i, U_i)$ and activation weights w_k of DRA-EBRBS are shown in Table 5. Consequently, the rule set is Δ_λ={R_1, R_2} and the function is $C(\Delta_\lambda)$=1. Because the upper bound of $C(\Delta_\lambda)$ is 1.0, R_1 and R_2 are selected as activated rules for DRA-EBRBS.

Table 5. Individual matching degrees , and activation weights in DRA-EBRBS

	$S^k(x_1, U_1)$	$S^k(x_2, U_2)$	$S^k(x_3, U_3)$	$S^k(x_4, U_4)$	$S^k(x_5, U_5)$	w_k
R_1	0.893	0.371	0.456	0.408	0.070	0.981
R_2	0.323	0.183	0.394	0.034	0.135	0.019
R_3	0.156	0.264	0.689	0.004	0.034	0.000

According to ***Step 2*** of the EBRB inference scheme shown in Section 3.2, the belief distribution of consequent attribute in R_1 and R_2 should be integrated using the analytical ER algorithm. Finally, the integrated belief distribution is therefore $\{(D_1, 0), (D_2, 0), (D_3, 1), (D_4, 0)\}$ and the output class is D_3 or *Middle*.

Continuing with the real scenario that the dataset of user knowledge modeling with 5-fold distribution optimally balanced cross validation are used to test O-EBRBS, DRA-EBRBS, and Micro-EBRBS, the accuracy of the three EBRBSs can be calculated though the above EBRB inference scheme and they are shown in Fig. 8. From Fig. 8, it is clear that the accuracy of DRA-EBRBS is higher than O-EBRBS and Micro-EBRBS which handing classification problem of user knowledge modeling.

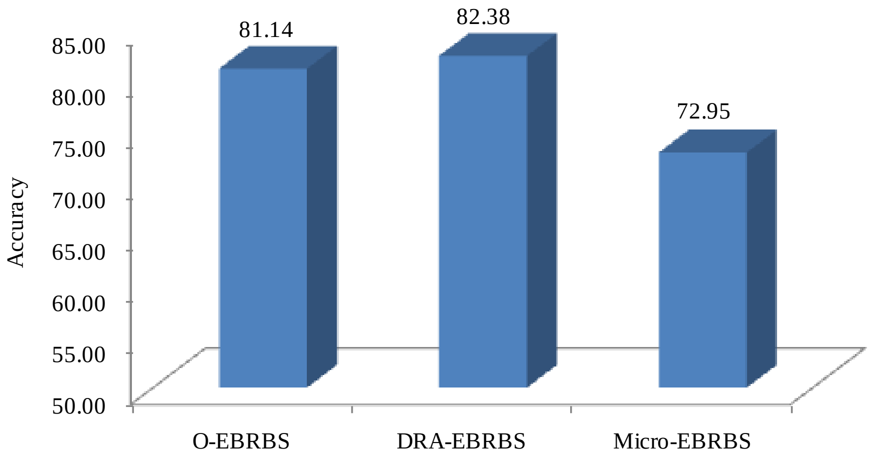


Fig. 8. Accuracy of O-EBRBS, DRA-EBRBS, and Micro-EBRBS

In conclusion, according to the real scenario of the EBRB generation process on user knowledge modeling, the number of extended belief rules in O-EBRBS and DRA-EBRBS is 322, which is equal to the number of training data. Owing to rule reduction, Micro-EBRBS only has 110 extended belief rules, which is far less than O-EBRBS and DRA-EBRBS. Correspondingly, according to the real scenario of the EBRB inference process on user knowledge modeling, the priority order of accuracies for the three EBRBSs is DRA-EBRBS > O-EBRBS > Micro-EBRBS.

5.3. Comparison of O-EBRBS, DRA-EBRBS, and Micro-EBRBS on the Performance

This subsection aims at describing the different performances among O-EBRBS, DRA-EBRBS, and Micro-EBRBS based on twenty classification datasets shown in Table 1. According to the EBRB generation scheme and the EBRB inference scheme, the results of O-EBRBS, DRA-EBRBS, and Micro-EBRBS can be obtained. Table 6 shows the mean and standard deviation in the terms of average accuracy, number of rules, rule generation time, inference time, and total time for the three EBRBSs.

Table 6. Comparison of O-EBRBS, DRA-EBRBS, and Micro-EBRBS in terms of mean and standard deviation

Classifier	Accuracy (%)	Number of rules	EBRB generation time (ms)	EBRB Inference time (ms)	Number of failed
O-EBRBS	79.79+15.76	602.25+521.90	1357.35+2397.16	500.90+950.81	2.50+6.45
DRA-EBRBS	81.65+14.71	602.25+521.90	1469.75+2595.11	6258.10+11970.73	0.00+0.00
Micro-EBRBS	77.02+15.39	275.91+358.17	287.20+517.68	180.10+244.29	2.60+6.49

From Table 6, three main differences of the three EBRBSs can be summarized as follows:

(1) Comparing to O-EBRBS and Micro-EBRBS, a remarkable advantage of DRA-EBRBS is that it has a higher accuracy for classifying twenty classification datasets. This is because no matter what input data are provided for DRA-EBRBS, the consistent activated rules can be determined by using the dynamic rule activation method in the EBRB inference scheme, which does not involve in O-EBRBS and Micro-EBRBS.

(2) Comparing to O-EBRBS and DRA-EBRBS, a notable merit of Micro-EBRBS is that it has a smaller number of rules, EBRB generation time, EBRB inference time, and total time obtained from twenty classification datasets. This is because Micro-EBRBS can combine multiple extended belief rules into a new rule so that the extended belief rules which make no sense should not be imposed on the EBRB generation scheme and EBRB inference scheme.

(3) Comparing to DRA-EBRBS and Micro-EBRBS, O-EBRBS is a “neutral” classifier in terms of accuracy, *i.e.*, accuracy rank is 81.65% (DRA-EBRBS) > 79.79% (O-EBRBS) > 77.02% (Micro-EBRBS), and computing time, *i.e.*, total time rank is 467.30ms (Micro-EBRBS) < 1858.25ms (O-EBRBS) < 7727.85ms (DRA-EBRBS). This is mainly attributed to lack of dynamic rule activation and rule reduction in the EBRB generation scheme and EBRB inference scheme.

In order to further compare O-EBRBS, DRA-EBRBS, and Micro-EBRBS, the Aligned Friedman test is applied to determine whether significant differences can be found in the results of the three EBRBSs. Table 7 shows the Aligned Friedman statistic for accuracy and total time and their corresponding critical values based on a level of significance of α=0.1. It is clear from Table 7 that two Aligned Friedman statistics are greater than their critical values, thus there are significant differences with α=0.1 for accuracy and total time. For example, the statistic value of accuracy 16.8307 is greater than the critical value 4.6052, leading to rejecting hypothesis, which means that the accuracy of the three EBRBSs has significant differences.

Table 7. Aligned Friedman tests to compare three EBRBSs (α=0.1)

Indicator	Statistic value	Critical value	Hypothesis
Accuracy	16.8307	4.6052	Rejected
Total time	30.2918	4.6052	Rejected

In order to further compare three EBRBSs with standard machine learning algorithms [53]-[56], KNN, Naïve Bayes (NB), DT, SVM, Artificial Neural Network (ANN), and Chi-FRBCS are introduced to perform a comparative analysis. It is worth noting that all these machine learning algorithms are basic algorithms without using ensemble learning to improve their performance. Table 8 shows the accuracy and rank of twenty classification datasets.

From Table 8, it can be found that the O-EBRBS obtains the best accuracy in Autos and Penbased, respectively, and they are 79.25% and 96.73%. Although the dynamic rule activation method fails to improve the accuracy of the O-EBRBS in Autos and Penbased, the DRA-EBRBS obtains the best accuracy in Glass, Tae, Vowel, and Wine, respectively, and they are 75.23%, 64.24%, 99.09%, and 98.31%. This means that the dynamic rule activation method can be regarded as a useful method to improve the general accuracy of EBRBSs. Comparatively, the Micro-EBRBS always obtain lower accuracies than the O-EBRBS and DRA-EBRBS. In the comparison of six machine learning algorithms, the rank is ANN (2.825) > DRA-EBRBS (3.900) > DT (4.525) = O-EBRBS (4.525) > NB (4.950) > KNN (5.075) > Micro-EBRBS (6.000) SVM (6.125) > > Chi-FRBCS (7.050), indicating that the EBRBSs can produce satisfactory results comparing to standard machine learning algorithms.

Table 8. Accuracy and rank of six machine learning algorithms and three EBRBSs

Classifier	KNN	NB	DT	SVM	ANN	Chi-FRBCS	O-EBRBS	DRA-EBRBS	Micro-EBRBS
Autos	60.38(7)	59.75(8)	76.73(4.5)	30.19(9)	76.73(4.5)	72.96 (6)	79.25 (1)	78.62 (2)	77.36 (3)
Car	86.40(5)	92.36(3)	91.84(4)	96.59(2)	98.73(1)	70.02(7.5)	70.02(7.5)	70.02(7.5)	70.02(7.5)
Contraceptive	49.49(5)	49.08(6.5)	52.95(3)	56.75(1)	54.24(2)	47.93(8)	52.14(4)	45.62(9)	49.08(6.5)
Ecoli	80.95(2)	80.36(3)	81.25(1)	42.56(9)	79.17(4.5)	76.49(6.5)	76.49(6.5)	79.17(4.5)	73.81(8)
Flare	75.14(2)	72.23(6)	73.26(4.5)	75.98(1)	73.26(4.5)	61.07(9)	66.89(7)	73.36(3)	66.79(8)
Glass	49.53(9)	67.76(5)	68.22(4)	65.89(6.5)	69.63(2)	50.00(8)	69.16(3)	75.23(1)	65.89(6.5)
Iris	96.00(4)	96.67(2)	94.67(6)	98.00(1)	96.00(4)	80.00(9)	96.00(4)	94.67(6.5)	90.00(8)
Knowledge	89.08(3)	85.61(4)	93.80(2)	81.14(6.5)	94.79(1)	26.30(9)	81.14(6.5)	82.38(5)	72.95(8)
Lymphography	82.43(2.5)	81.76(5)	75.00(9)	79.73(7.5)	83.78(1)	83.11(4)	79.73(7.5)	82.43(2.5)	80.41(6)
Nursery	88.66(7)	90.20(5)	89.51(6)	87.58(8)	95.68(1)	90.66(3.5)	90.66(3.5)	87.19(9)	90.74(2)
Penbased	84.73(8)	95.00(5)	88.91(7)	11.36(9)	93.45(6)	95.55(4)	96.73(1)	96.64(2)	96.36(3)
Seeds	90.48(5)	92.86(2)	91.90(3)	89.52(7)	94.76(1)	85.24(8.5)	90.95(4)	90.00(6)	85.24(8.5)
Segment	80.48(8)	94.85(4)	96.45(1.5)	65.19(9)	96.45(1.5)	81.82(7)	92.03(5)	96.28(3)	87.36(6)
Tae	54.30(3)	53.64(6)	49.67(8)	54.30(3)	53.64(6)	45.70(9)	54.30(3)	64.24(1)	53.64(6)
Thyroid	95.00(2)	92.50(4.5)	98.61(1)	92.50(4.5)	93.61(3)	88.06(9)	88.19(7.5)	92.36(6)	88.19(7.5)
Vehicle	45.27(8)	69.03(5)	73.88(2)	30.85(9)	82.39(1)	65.49(7)	70.33(4)	70.92(3)	68.79(6)
Vowel	70.71(8)	78.59(7)	80.71(6)	89.90(3)	86.97(4)	63.23(9)	97.68(2)	99.09(1)	82.83(5)
Wine	97.75(2.5)	96.07(6)	92.70(7)	44.38(9)	97.75(2.5)	93.26(8)	97.19(4.5)	98.31(1)	97.19(4.5)
Yeast	58.36(3.5)	58.36(3.5)	56.27(5)	42.18(9)	59.70(1)	43.94(8)	49.87(6)	59.43(2)	46.70(7)
Zoo	93.07(7)	90.10(8.5)	94.06(6)	90.10(8.5)	95.05(5)	98.02(1)	97.03(3)	97.03(3)	97.03(3)
Rank	5.075	4.950	4.525	6.125	2.825	7.050	4.525	3.900	6.000

In summary, we can observe that DRA-EBRBS is proven to be an "accuracy first" classifier, Micro-EBRBS is an "efficiency first" classifier, and O-EBRBS is a "neutral" classifier. All of them can comprehensively represent different stands of EBRBSs and ensure an extensive investigation of the influence of decomposition strategies and overlap functions on EBRBSs.

5.4. Influence of Decomposition Strategy and Overlap Function on the Performance of EBRBSs

This subsection aims at analyzing the influence of decomposition strategies and overlap functions on the performance of O-EBRBS, DRA-EBRBS, and Micro-EBRBS. In order to introduce the process of using the three EBRBSs with OVO strategy and five overlap functions to handle classification datasets, the details of using the OVO strategy shown in Section 4.2 and the five overlap functions shown in Section 4.3 are provided when handling classification problem Knowledge as follows:

(1) EBRBSs with OVO strategy to classify input data.

Firstly, the classification problem Knowledge, which includes four classes namely *Very Low* (D_1), *Low* (D_2), *Middle* (D_3), and *High* (D_4), should be divided into six sub-datasets with two-class, namely{D_1, D_2}, {D_1, D_3}, {D_1, D_4}, {D_2, D_3}, {D_2, D_4}, and {D_3, D_4}. Accordingly, six binary EBRBSs can be constructed on the basis of six sub-datasets using the EBRB generation scheme shown in Section 3. Furthermore, suppose that the output belief distribution of the six binary EBRBSs to classify input data $\boldsymbol{x}$ is {(D_1, 0.6), (D_2, 0.4)}, {D_1, 0.8), (D_3, 0.2)}, {D_1, 0.7), (D_4, 0.3)}, {(D_2, 0.7), (D_3, 0.3)}, {(D_2, 0.8), (D_4, 0.2)}, and {(D_3, 0.6), (D_4, 0.4)}, respectively. Hence, the matrix $\hat{\boldsymbol{\beta}}(\boldsymbol{x})$ is expressed as follows:

$$\hat{\boldsymbol{\beta}}(\boldsymbol{x})=\begin{pmatrix} - & 0.6 & 0.8 & 0.7 \\ 0.4 & - & 0.7 & 0.8 \\ 0.2 & 0.3 & - & 0.6 \\ 0.3 & 0.2 & 0.4 & - \end{pmatrix} \tag{42}$$

For the VS function, $s_{i,j}$ (i, j=1,…, 4) should be calculated based on Eq. (26) shown in Section 4.2 and they are $s_{1,2} = s_{1,3} = s_{1,4} = s_{2,3} = s_{2,4} = s_{3,4} = 1$ and $s_{2,1} = s_{3,1} = s_{4,1} = s_{3,2} = s_{4,2} = s_{4,3} = 0$. Hence, the final output class is D_1 because of n = arg max $\{s_{1,2} + s_{1,3} + s_{1,4} = 3, s_{2,1} + s_{2,3} + s_{2,4} = 2, s_{3,1} + s_{3,2} + s_{3,4} = 1, s_{4,1} + s_{4,2} + s_{4,3} = 0\} = 1$ according to Eq. (25) shown in Section 4.2.

For the WV function, the final output class is D_1 because of n=arg max $\{\hat{\beta}_{1,2}+\hat{\beta}_{1,3}+\hat{\beta}_{1,4}= 2.1, \hat{\beta}_{2,1}+\hat{\beta}_{2,3}+\hat{\beta}_{2,4}= 1.9, \hat{\beta}_{3,1}+\hat{\beta}_{3,2}+\hat{\beta}_{3,4} = 1.1, \hat{\beta}_{4,1}+\hat{\beta}_{4,2}+\hat{\beta}_{4,3}= 0.9\} = 1$ according to Eq. (27) shown in Section 4.2.

For the WWV function, $s_{i,j}$ (i, j=1,…, 4) should be calculated based on Eq. (28) shown in Section 4.2 and they are $s_{1,2} = 0.6$, $s_{1,3} = 0.8$, $s_{1,4} = 0.7$, $s_{2,3} = 0.7$, $s_{2,4} = 0.8$, $s_{3,4} = 0.6$, and $s_{2,1} = s_{3,1} = s_{4,1} = s_{3,2} = s_{4,2} = s_{4,3} = 0$. Hence, the final output class is D_1 because of n = arg max $\{s_{1,2} + s_{1,3} + s_{1,4} = 2.1, s_{2,1} + s_{2,3} + s_{2,4} = 1.5, s_{3,1} + s_{3,2} + s_{3,4} = 0.6, s_{4,1} + s_{4,2} + s_{4,3} = 0\} = 1$ according to Eq. (29) shown in Section 4.2.

For the NC function, $t_{i,j}$ (i, j=1,…, 4) should be calculated based on Eq. (30) shown in Section 4.2 and they are $t_{1,2} = 0.2$, $t_{1,3} = 0.6$, $t_{1,4} = 0.4$, $t_{2,3} = 0.4$, $t_{2,4} = 0.6$, $t_{3,4} = 0.2$,and $t_{2,1} = t_{3,1} = t_{4,1} = t_{3,2} = t_{4,2} = t_{4,3} = 0$. Hence, the final output class is D_1 because of n = arg max{1 - max$\{t_{2,1}, t_{3,1}, t_{4,1}\}$ = 1 - max{0, 0, 0} = 1, 1 - max$\{t_{1,2}, t_{3,2}, t_{4,2}\}$ = 1 - max{0.2, 0, 0} = 0.8, 1 - max$\{t_{1,3}, t_{2,3}, t_{4,3}\}$ = 1 - max{0.6, 0.4, 0} = 0.4, 1 - max$\{t_{1,4}, t_{2,4}, t_{3,4}\}$ = 1 - max{0.4, 0.6, 0.2} = 0.4} = 1 according to Eq. (31) shown in Section 4.2.

For the LC function, suppose that the number of input-output data pairs related to D_1, D_2, D_3, and D_4, is $N_1 = 3$, $N_2 = 3$, $N_3 = 3$, and $N_4 = 3$, respectively. The conflict degrees $C_{i,j}$ (i, j=1,…, 4) should be calculated and they are $C_{2,1} = C_{1,2} = \min\{\hat{\beta}_{1,2} =0.6, \hat{\beta}_{2,1} = 0.4\} = 0.4$, $C_{3,1} = C_{1,3} = \min\{\hat{\beta}_{1,3} = 0.8, \hat{\beta}_{3,1} = 0.2\} = 0.2$, $C_{4,1} = C_{1,4} = \min\{\hat{\beta}_{1,4} = 0.7, \hat{\beta}_{4,1} = 0.3\} = 0.3$, $C_{3,2} = C_{2,3} = \min\{\hat{\beta}_{2,3} = 0.7, \hat{\beta}_{3,2} = 0.3\} = 0.3$, $C_{4,2} = C_{2,4} = \min\{\hat{\beta}_{2,4} = 0.8, \hat{\beta}_{4,2} = 0.2\} = 0.2$, and , $C_{4,3} = C_{3,4} = \min\{\hat{\beta}_{3,4} = 0.6, \hat{\beta}_{4,3} = 0.4\} = 0.4$. The strict preferences $P_{i,j}$ (i, j=1,…, 4) should be calculated and they are $P_{1,2} = \hat{\beta}_{1,2} - C_{1,2} = 0.2$, $P_{1,3} = \hat{\beta}_{1,3} - C_{1,3} = 0.6$, $P_{1,4} = \hat{\beta}_{1,4} - C_{1,4} = 0.4$, $P_{2,1} = \hat{\beta}_{2,1} - C_{2,1} = 0$, $P_{2,3} = \hat{\beta}_{2,3} - C_{2,3} = 0.4$, $P_{2,4} = \hat{\beta}_{2,4} - C_{2,4} = 0.6$, $P_{3,1} = \hat{\beta}_{3,1} - C_{3,1} = 0$, $P_{3,2} = \hat{\beta}_{3,2} - C_{3,2} = 0$, $P_{3,4} = \hat{\beta}_{3,4} - C_{3,4} = 0.2$, $P_{4,1} = \hat{\beta}_{4,1} - C_{4,1} = 0$, $P_{4,2} = \hat{\beta}_{4,2} - C_{4,2} = 0$, $P_{4,3} = \hat{\beta}_{4,3} - C_{4,3} = 0$. The ignorance degrees $I_{i,j}$ (i, j=1,…, 4) should be calculated and they are $I_{1,2} = I_{2,1} = 1 - \max\{\hat{\beta}_{1,2} = 0.6, \hat{\beta}_{2,1} = 0.4\} = 0.4$, $I_{1,3} = I_{3,1} = 1 - \max\{\hat{\beta}_{1,3} = 0.8, \hat{\beta}_{3,1} = 0.2\} = 0.2$, $I_{1,4} = I_{4,1} = 1 - \max\{\hat{\beta}_{1,4} = 0.7, \hat{\beta}_{4,1} = 0.3\} = 0.3$, $I_{2,3} = I_{3,2} = 1 - \max\{\hat{\beta}_{2,3} =0.7, \hat{\beta}_{3,2} =0.3\} =0.3$, $I_{2,4} = I_{4,2} = 1 - \max\{\hat{\beta}_{2,4} =0.8, \hat{\beta}_{4,2} =0.2\} =0.2$, and $I_{3,4} = I_{4,3} = 1 - \max\{\hat{\beta}_{3,4} =0.6, \hat{\beta}_{4,3} =0.4\} =0.4$. Hence, the final output class is D_1 because of n = arg max$\{P_{1,2} + P_{1,3} + P_{1,4} + (C_{1,2} + C_{1,3}+ C_{1,4}) / 2 + (N_1 / (N_1 + N_2) \times I_{1,2}+ N_1 / (N_1 + N_3) \times I_{1,3}+ N_1 / (N_1 + N_4) \times I_{1,4}) = 2.1, P_{2,1} + P_{2,3} + P_{2,4} + (C_{2,1} + C_{2,3} + C_{2,4}) / 2 + (N_2 / (N_2 + N_1) \times I_{2,1} + N_2 / (N_2 + N_3) \times I_{2,3} + N_2 / (N_2 + N_4) \times I_{2,4}) = 1.9, P_{3,1} + P_{3,2} + P_{3,4} + (C_{3,1} + C_{3,2} + C_{3,4}) / 2 + (N_3 / (N_3 + N_1) \times I_{3,1} + N_3 / (N_3 + N_2) \times I_{3,2}+ N_3 / (N_3 + N_4) \times I_{3,4}) = 1.1, P_{4,1} + P_{4,2} + P_{4,3} + (C_{4,1} + C_{4,2} + C_{4,3}) / 2 + (N_4 / (N_4 + N_1) \times I_{4,1} + N_4 / (N_4 + N_2) \times I_{4,2} + N_4 / (N_4 + N_3) \times I_{4,3}) = 0.9\} = 1$ according to Eq. (32) shown in Section 4.2.

(2) The EBRBSs with five overlap functions to classify input data.

Continuing with the individual matching degrees shown in Table 4 as an example, the five overlap functions, including PR, MI, SP, GM, and SI functions, are used to calculate the activation weight of each extended belief rule. Table 9 shows the activation weight of R_k, R_t, and R_l obtained from Eqs. (33) to (38). It is clear from Table 9 that R_k and R_t should be activated with different activation weights when using the SP, PR, MI, GM, SI functions, respectively, where the maximum difference between the activation weights of R_k and R_t can be found in the SP function and the minimum one in the SI function, which reflect the same relationship as shown in Fig. 6 at Section 4.3.

Table 9. Activation weight of three rules when using PR, MI, SP, GM, and SI functions

R_k	$S^k(x_1, U_1)$	$S^k(x_2, U_2)$	$S^k(x_3, U_3)$	$S^k(x_4, U_4)$	$S^k(x_5, U_5)$	θ_k	w_k				
							SP	PR	MI	GM	SI
R_1	0.972	0.780	0.822	0.799	0.515	0.987	0.8832	0.7368	0.5615	0.5462	0.5013
R_2	0.754	0.654	0.792	0.429	0.606	0.789	0.1107	0.2332	0.3739	0.3628	0.3880
R_3	0.628	0.717	0.911	0.250	0.429	0.234	0.0062	0.0300	0.0646	0.0910	0.1107

According to the above process of using OVO strategy and five overlap functions to enhance three EBRBSs, the results of O-EBRBS, DRA-EBRBS, and Micro-EBRBS for the twenty classification datasets shown in Table 1 can be obtained. Tables 10, 13, and 15 show the average accuracy of EBRBSs, where these results are obtained from each baseline EBRBSs along with the OVO strategy with five aggregation functions, denoted as OVO^{VS}, OVO^{WV}, OVO^{WWV}, OVO^{NC}, and OVO^{LC}, by using five overlap functions to model the conjunctive relationship among antecedent attributes in enhanced EBRBSs. As shown in Tables 10, 13, and 15, the best average accuracy obtained from OVO strategy is highlighted in underline and the best one obtained from five overlap functions is highlighted in boldface. Additionally, in order to investigate significant differences among the results of each OVO-related strategy or overlap function in a given EBRBS, the Aligned Friedman and Holm tests are carried out to obtain the average rank of accuracies and the adjusted *p* value, and the corresponding results can be found in Tables 11-12, 14-15, and 17-18, in which the best rank of average accuracies is highlighted in boldface and the significant difference based on the adjusted *p* value is highlighted in underline.

Table 10. Average accuracy and standard deviation of O-EBRBS

	PR	SP	MI	GM	SI
Baseline	79.79+15.76	**81.67+14.88**	64.38+20.95	63.33+20.08	56.67+21.93
OVO^{VS}	79.82+15.76	**81.71+14.89**	64.42+20.97	<u>63.36+20.07</u>	56.73+21.98
OVO^{WV}	<u>79.88+15.76</u>	**<u>81.82+14.77</u>**	<u>64.46+20.97</u>	<u>63.36+20.07</u>	56.73+21.98
OVO^{WWV}	79.82+15.76	**81.71+14.89**	64.45+20.96	<u>63.36+20.07</u>	56.73+21.98
OVO^{NC}	79.82+15.76	**81.71+14.89**	64.42+20.97	<u>63.36+20.07</u>	56.73+21.98
OVO^{LC}	79.23+16.53	**81.86+15.04**	64.42+21.56	63.10+20.61	<u>56.92+22.24</u>

Table 11. Aligned Friedman and Holm tests to compare overlap functions on O-EBRBS

	Average rank of accuracy					Adjusted *p* value (α = 0.1)			
	PR	SP	MI	GM	SI	SP	MI	GM	SI
Baseline	24.525	**23.675**	60.050	65.825	78.425	0.9262	<u>0.0001</u>	<u>0.0000</u>	<u>0.0000</u>
OVO^{VS}	24.575	**23.675**	60.050	65.925	78.275	0.9219	<u>0.0001</u>	<u>0.0000</u>	<u>0.0000</u>
OVO^{WV}	24.575	**23.525**	60.000	66.025	78.375	0.9089	<u>0.0001</u>	<u>0.0000</u>	<u>0.0000</u>
OVO^{WWV}	24.575	**23.675**	60.000	65.975	78.275	0.9219	<u>0.0001</u>	<u>0.0000</u>	<u>0.0000</u>
OVO^{NC}	24.575	**23.675**	60.050	65.925	78.275	0.9219	<u>0.0001</u>	<u>0.0000</u>	<u>0.0000</u>
OVO^{LC}	25.675	**24.100**	59.475	65.500	77.750	0.8637	<u>0.0002</u>	<u>0.0000</u>	<u>0.0000</u>

Table 12. Aligned Friedman and Holm tests to compare decomposition strategies on O-EBRBS

	Average rank of accuracy						Adjusted p value (α = 0.1)				
	Baseline	OVO^{VS}	OVO^{WV}	OVO^{WWV}	OVO^{NC}	OVO^{LC}	OVO^{VS}	OVO^{WV}	OVO^{WWV}	OVO^{NC}	OVO^{LC}
PR	64.725	58.975	**50.425**	58.975	58.975	70.925	0.6012	0.1936	0.6012	0.6012	0.5730
SP	68.875	64.150	**50.375**	64.150	64.150	51.300	0.6675	0.0926	0.6675	0.6675	0.1101
MI	70.500	60.950	**53.275**	56.600	60.950	60.725	0.3853	0.1174	0.2064	0.3853	0.3742
GM	68.325	**56.375**	**56.375**	**56.375**	**56.375**	69.175	0.2773	0.2773	0.2773	0.2773	0.9384
SI	65.325	**56.625**	**56.625**	**56.625**	56.625	71.175	0.4290	0.4290	0.4290	0.4290	0.5949

Tables 10, 11, and 12 show the performance influences of OVO^{VS}, OVO^{WV}, OVO^{WWV}, OVO^{NC}, and OVO^{LC} strategies and PR, SP, MI, GM, and SI functions on O-EBRBS and these influences can be summarized as follows:

(1) As we can observe in Table 10, OVO^{WV} strategy has a better improvement in O-EBRBS than other OVO-related strategies. For the overlap functions, SP function has the best accuracy in comparison with other overlap functions. More importantly, from Tables 11 and 12, the average rank of the accuracies obtained from twenty datasets also shows that OVO^{WV} strategy and SP function can achieve the best classification performance from other OVO-related strategies or overlap functions.

(2) Table 10 shows that the accuracies obtained from SP function are significantly better than other overlap functions in O-EBRBS. This situation is confirmed by the Aligned Friedman and Holm tests in Table 11, in which the significant differences are only found in MI, GM, and SI functions. The reason for these influences is that, as we can observe in Fig. 6, SP function can produce a smaller $O(S_1^k,...,S_M^k)$ so that it is more powerful to distinguish consistent activated rules.

(3) From Table 10, the accuracies obtained from OVO^{VS}, OVO^{WV}, OVO^{WWV}, and OVO^{NC} strategies are obviously greater than those of the baselines of O-EBRBS. Meanwhile, it is clear from Table 12 that OVO^{WV} strategy has much better rank of average accuracy and there are significant differences in comparison with the baselines of O-EBRBS. The reason for this influence is that OVO-related strategies make full use of binary EBRBSs to produce final output class.

Table 13. Average accuracy and standard deviation of DRA-EBRBS

	PR	SP	MI	GM	SI
Baseline	**81.65+14.71**	81.50+14.78	71.47+21.71	62.47+22.29	56.33+22.62
OVO^{VS}	79.84+16.28	**82.06+15.41**	68.27+21.97	57.52+21.80	54.31+22.67
OVO^{WV}	80.36+15.67	**82.05+15.40**	69.00+22.33	59.23+20.72	55.18+21.89
OVO^{WWV}	79.93+16.14	**82.06+15.44**	68.61+21.99	57.82+21.56	54.33+22.66
OVO^{NC}	79.96+16.14	**82.08+15.42**	68.76+22.22	58.18+21.17	54.33+22.64
OVO^{LC}	79.92+15.93	**81.96+15.36**	68.79+22.35	58.99+20.81	55.13+21.93

Table 14. Aligned Friedman and Holm tests to compare overlap functions on DRA-EBRBS

	Average rank of accuracy					Adjusted p value (α = 0.1)			
	PR	SP	MI	GM	SI	SP	MI	GM	SI
Baseline	**26.700**	28.025	47.900	69.825	80.050	0.8852	0.0208	0.0000	0.0000
OVO^{VS}	**28.350**	29.950	48.500	70.700	75.000	0.8616	0.0281	0.0000	0.0000
OVO^{WV}	**27.850**	30.550	48.950	69.750	75.400	0.7685	0.0215	0.0000	0.0000
OVO^{WWV}	**28.500**	30.250	48.050	70.400	75.300	0.8487	0.0331	0.0000	0.0000
OVO^{NC}	**28.475**	30.300	48.075	70.400	75.250	0.8423	0.0326	0.0000	0.0000
OVO^{LC}	**28.300**	30.375	48.625	69.950	75.250	0.8211	0.0267	0.0000	0.0000

Table 15. Aligned Friedman and Holm tests to compare decomposition strategies on DRA-EBRBS

	Average rank of accuracy						Adjusted p value (α = 0.1)				
	Baseline	OVO^{VS}	OVO^{WV}	OVO^{WWV}	OVO^{NC}	OVO^{LC}	OVO^{VS}	OVO^{WV}	OVO^{WWV}	OVO^{NC}	OVO^{LC}
PR	**32.525**	69.375	54.325	68.600	65.825	72.350	0.0008	0.0475	0.0010	0.0025	0.0003
SP	61.000	60.200	60.650	60.125	**59.600**	61.425	0.9420	0.9746	0.9366	0.8987	0.9692
MI	**34.200**	72.275	57.675	68.175	68.300	62.375	0.0005	0.0328	0.0020	0.0019	0.0104
GM	**27.725**	73.850	55.375	71.075	71.100	63.875	0.0001	0.0119	0.0003	0.0003	0.0020
SI	**37.775**	70.100	56.325	69.950	69.950	58.900	0.0033	0.0917	0.0034	0.0034	0.0548

Tables 13, 14, and 15 show the performance influences of OVO^{VS}, OVO^{WV}, OVO^{WWV}, OVO^{NC}, and OVO^{LC} strategies and PR, SP, MI, GM, and SI functions on DRA-EBRBS and these influences can be summarized as follows:

(1) Looking at the accuracy shown in Table 13, we can observe that both of SP and PR functions are able to obtain the best accuracy for DRA-EBRBS. For example, SP function outperforms RP function in five kinds of OVO strategies. For the influence of decomposition strategy on DRA-EBRBS, OVO strategy fails to get the better accuracy on all kinds of overlap functions.

(2) As shown in Table 14, the average rank of accuracy obtained from SP or PR function is is better than other functions on DRA-EBRBS. The reason for this difference is that, as described in Section 3.3, the dynamic rule activation method is able to re-adjust individual matching degrees to activate consistent rules. For example, when the value of λ is 2 and the value of S_1^k and S_2^k is 0.2 and 0.3, the produced value of MI function is $O(S_1^k, S_2^k)$ =0.04, which is even smaller than that of PR function.

(3) Taking a look at Table 13, the five OVO-related strategies have negative influence on DRA-EBRBS. From Table 15, the results of Aligned Friedman and Holm tests show that there are significant differences for PR, SP, GM, and SI functions when using the five OVO-related strategies to enhance DRA-EBRBS. The reason is that the dynamic rule activation method should depend on the belief distribution related to all classes in determining consist activated rules, but OVO strategy makes the belief distributions only to consider two classes.

Table 16. Average accuracy and standard deviation of Micro-EBRBS

	PR	SP	MI	GM	SI
Baseline	77.02+15.39	**78.90+14.21**	62.62+20.57	61.17+20.73	55.77+22.48
OVO^{VS}	77.19+15.33	**79.18+14.31**	62.76+20.73	61.89+20.26	55.85+22.40
OVO^{WV}	77.87+15.61	**79.52+14.37**	63.41+20.72	61.97+20.40	56.37+21.80
OVO^{WWV}	77.35+15.39	**79.23+14.33**	62.82+20.72	61.90+20.23	55.94+22.30
OVO^{NC}	77.25+15.32	**79.25+14.35**	62.74+20.73	61.89+20.25	55.87+22.40
OVO^{LC}	77.11+16.60	**79.57+15.03**	63.19+21.45	62.21+20.50	56.09+21.96

Table 17. Aligned Friedman and Holm tests to compare overlap functions on Micro-EBRBS

	Average rank of accuracy					Adjusted p value (α = 0.1)			
	PR	SP	MI	GM	SI	SP	MI	GM	SI
Baseline	25.200	**24.575**	59.400	67.025	76.300	0.9457	0.0002	0.0000	0.0000
OVO^{VS}	25.475	**24.100**	60.400	65.950	76.575	0.8809	0.0001	0.0000	0.0000
OVO^{WV}	25.550	**24.525**	59.325	66.025	77.075	0.9110	0.0002	0.0000	0.0000
OVO^{WWV}	25.375	**24.150**	60.450	66.000	76.525	0.8938	0.0001	0.0000	0.0000
OVO^{NC}	25.375	**24.100**	60.550	66.000	76.475	0.8895	0.0001	0.0000	0.0000
OVO^{LC}	26.400	**24.550**	59.600	65.025	76.925	0.8402	0.0003	0.0000	0.0000

Table 18. Aligned Friedman and Holm tests to compare decomposition strategies on Micro-EBRBS

	Average rank of accuracy						Adjusted p value (α = 0.1)				
	Baseline	OVO^{VS}	OVO^{WV}	OVO^{WWV}	OVO^{NC}	OVO^{LC}	OVO^{VS}	OVO^{WV}	OVO^{WWV}	OVO^{NC}	OVO^{LC}
PR	69.775	65.575	**49.000**	56.750	64.050	57.850	0.7026	0.0589	0.2364	0.6028	0.2783
SP	76.125	65.125	53.525	60.750	59.375	**48.100**	0.3173	0.0399	0.1622	0.1278	0.0108
MI	71.450	64.050	**45.950**	66.775	67.500	47.275	0.5011	0.0204	0.6708	0.7195	0.0280
GM	72.750	61.375	**53.525**	61.425	60.275	53.650	0.3011	0.0805	0.3032	0.2568	0.0825
SI	63.475	62.625	**54.875**	59.875	57.050	65.100	0.9384	0.4343	0.7435	0.5592	0.8826

Tables 16, 17, and 18 show the performance influences of OVO^{VS}, OVO^{WV}, OVO^{WWV}, OVO^{NC}, and OVO^{LC} strategies and PR, SP, MI, GM, and SI functions on Micro-EBRBS and these influences can be summarized as follows:

(1) Table 16 shows that the influence of OVO^{VS}, OVO^{WV}, OVO^{WWV}, OVO^{NC}, and OVO^{LC} strategies and PR, SP, MI, GM, and SI functions is similar to that of O-EBRBS. In other words, Micro-EBRBS is also able to obtain the best average accuracy obtained from the twenty datasets when using SP function and OVO^{WV} strategy to improve the classification performance of Micro-EBRBS. All these situations can be confirmed in the average ranks of accuracies shown in Tables 11 and 12.

(2) From Table 17, the significant differences can be found in MI, GM, and SI functions with OVO^{VS}, OVO^{WV}, OVO^{WWV}, OVO^{NC}, and OVO^{LC} strategies. The reason for this influence is the same to O-EBRBS because, as shown in Fig. 5, Micro-EBRBS has the same inference scheme in the comparison with O-EBRBS, so that SP function can be a powerful function better than PR, MI, GM, and SI functions in Micro-EBRBS to distinguish consistent activated rules.

(3) Taking a look at Table 18, the accuracies obtained from OVO^{VS}, OVO^{WV}, OVO^{WWV}, OVO^{NC}, and OVO^{LC} strategies are difference from the baselines of Micro-EBRBS. The reason is that, for Micro-EBRBS, the rules mapped into division domains only include two kinds of classes because of the binary EBRBSs constructed by the five OVO-related strategies. Thus, each binary Micro-EBRBS has completely different rules comparing to the baselines and finally leads to different accuracies.

In summary, we can observe that the influences of overlap functions on O-EBRBS, DRA-EBRBS, and Micro-EBRBS are dependent on the capability of distinguishing consistent activated rules. Therefore, the best overlap function, which is able to improve the classification performance of EBRBSs, may be those can extract consistent activated rules for each test input data. Regarding decomposition strategies, the obtained accuracies and statistical analyses show that OVO^{VS}, OVO^{WV}, OVO^{WWV}, OVO^{NC}, and OVO^{LC} strategies are all able to improve the classification performance of O-EBRBS and Micro-EBRBS because these strategies make full use of each binary EBRBS to determine final output classes, especially for OVO^{WV} strategy .

5.5. Influence of Decomposition Strategy and Overlap Function on the Rule Base of EBRBSs

To further investigate the influences of OVO-related strategies and five overlap functions on O-EBRBS, DRA-EBRBS, and Micro-EBRBS, this subsection aims at showing these influences on the rule base of these EBRBSs. Table 19 presents the average number of rules, activated rules, and failed data for the three EBRBSs, respectively, in which the failed data means the number of tests where the EBRBS could not activate any rule to produce the final output class.

As shown in Table 19, in terms of PR, SP, MI, GM, and SI functions, O-EBRBS, DRA-EBRBS, and Micro-EBRBS have the same number of rules, such as 602.25 rules for the baseline of O-EBRBS and 275.81 rules for the baseline of Micro-EBRRS. Under the same condition, the OVO strategy has to raise the number of rules for three ERBBSs because binary EBRBSs are constructed by using sub datasets. For example, the training input-output data, which belongs to the *n*th (n=1,…, N) class, should be used to construct N-1 binary EBRBSs which are all related with the *n*th class.

Table 19. Average number of rules, activated rules, and failed data of EBRBSs in terms of mean and standard deviation

Function	Indicator	O-EBRBS		DRA-EBRBS		Micro-EBRBS	
		Baseline	OVO	Baseline	OVO	Baseline	OVO
PR	Rule	602.25+521.90	2807.96+3219.27	602.25+521.90	3106.89+3550.31	275.81+358.17	1220.69+1343.02
	Activated rule	338.36+411.79	1635.71+2391.74	106.63+166.30	1033.60+1540.14	175.99+362.34	745.04+1231.32
	Failed data	2.50+5.45	0.00+0.00	0.00+0.00	0.00+0.00	2.60+6.49	0.00+0.00
SP	Rule	602.26+521.90	2706.65+3107.40	602.25+521.90	3106.89+3550.31	275.81+358.17	1170.58+1307.13
	Activated rule	313.15+413.46	1519.24+2378.39	55.13+108.80	888.28+1390.25	165.98+362.15	698.76+1245.50
	Failed data	3.45+7.10	0.00+0.00	0.00+0.00	0.00+0.00	3.60+7.08	0.00
MI	Rule	602.25+521.90	2812.71+3225.03	602.25+521.90	3106.89+3550.31	275.81+358.17	1223.28+1345.15
	Activated rule	341.37+412.17	1649.50+2395.84	313.12+407.54	1821.82+2395.09	177.35+362.11	751.05+1229.95
	Failed data	2.45+6.44	0.00+0.00	0.00+0.00	0.00+0.00	2.55+6.48	0.00+0.00
GM	Rule	602.25+521.90	2812.71+3225.03	602.25+521.90	3106.89+3550.31	275.81+358.17	1223.28+1345.15
	Activated rule	341.37+412.17	1649.50+2395.84	377.25+418.83	2095.65+2531.98	177.35+362.11	751.05+1229.95
	Failed data	2.45+6.44	0.00+0.00	0.00+0.00	0.00+0.00	2.55+6.48	0.00+0.00
SI	Rule	602.25+521.90	2812.71+3225.03	602.25+521.90	3106.89+3550.31	275.81+358.17	1223.28+1345.15
	Activated rule	341.37+412.17	1649.50+2395.84	377.84+418.47	2097.50+2530.80	177.35+362.11	751.05+1229.95
	Failed data	2.45+6.44	0.00+0.00	0.00+0.00	0.00+0.00	2.55+6.48	0.00+0.00

For the number of activated rules, the OVO strategy is required to activate the same number of extended belief rules for O-EBRBS, DRA-EBRBS, and Micro-EBRBS, but is much more than that of the baseline of these three EBRBSs due to the fact that each binary EBRBS has to activate rules independently while classifying input data. Comparing to O-EBRBS, DRA-EBRBS and Micro-EBRBS usually activate fewer number of rules, in which DRA-EBRBS is attributed to the dynamic rule activation method which is able to activate consistent rules, and Micro-EBRBS is due to fewer number of rules in the rule base.

For the number of failed data, the baseline of O-EBRBS and Micro-EBRBS fails to activate rules in some of input data and their number of failed data is therefore greater than 0, but the OVO strategy can enhance the capability of these two EBRBSs to produce a final output class so that the enhanced EBRBSs are able to classify any given input data. Correspondingly, the baseline of DRA-EBRBS is able to avoid the situation of producing failed data. This is because the dynamic rule activation method not only is able to determine consistent activated rules by increasing the value the parameter λ, but also can activate all rules for any given input data by setting λ=0.

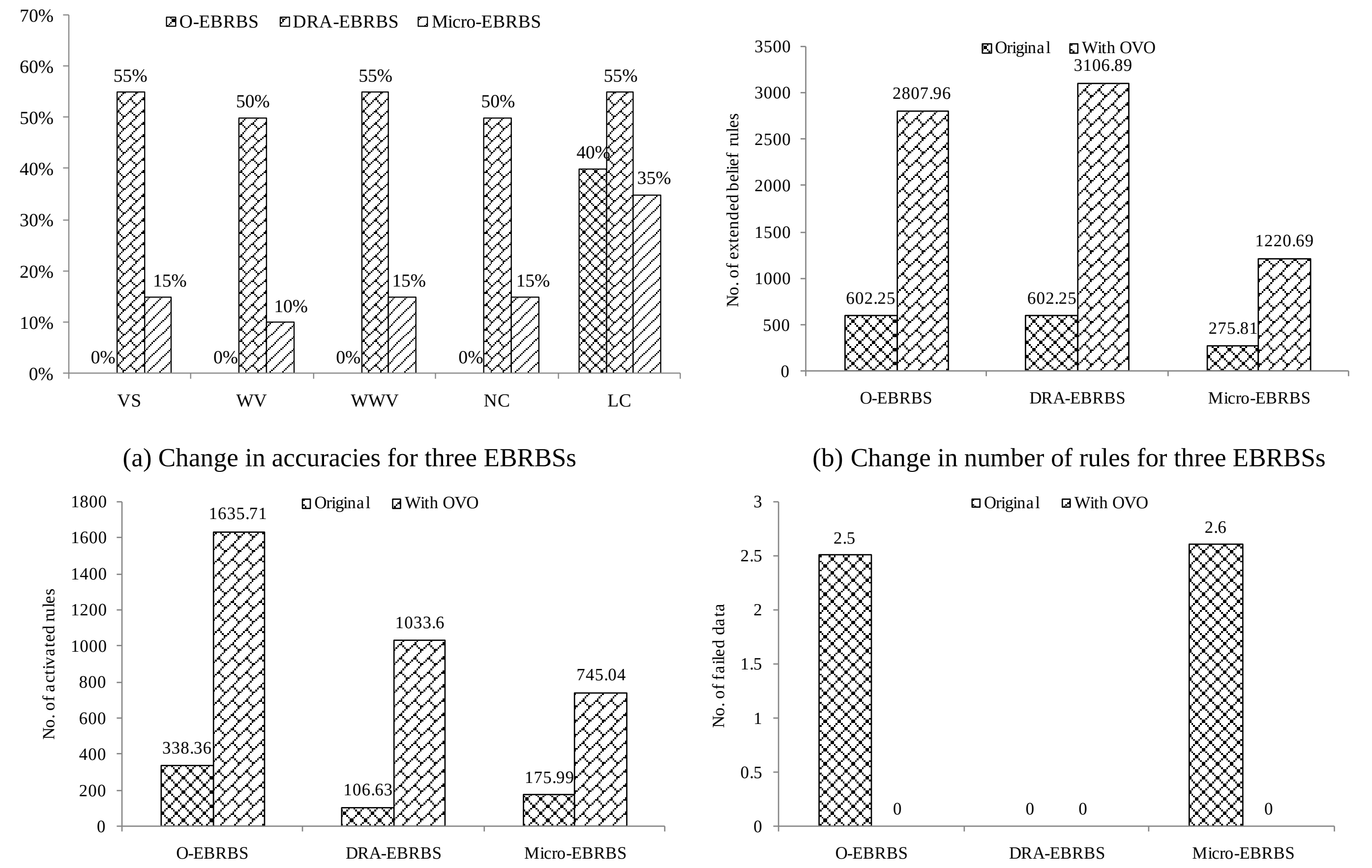
5.6. Discussions

After analyzing the performance of decomposition strategy and overlap function together with their impact on rule base, the OVO strategy and SP function have shown their advantages on enhanced O-EBRBS, DRA-EBRBS, and Micro-EBRBS. However, it still needs to discuss the reason why the improvement of the three EBRBS can be made. For this reason, this section is going to further investigate all the previously mentioned influences.

(1) Discussion of the influence of decomposition strategies on EBRBSs

The usage of OVO strategy affects not only the classification performance of EBRBSs, but also rule base. Fig. 9 provides the quantitative influences of OVO strategy on O-EBRBS, DRA-EBRBS, and Micro-EBRBS, in which Fig. 9(a) shows the percentage of the twenty datasets that the classification accuracy would be decreased after using the OVO strategy with five aggregation functions to enhance O-EBRBS, DRA-EBRBS, and Micro-EBRBS. In the case of O-EBRBS, the accuracies of twenty datasets do not decrease when using OVO^{VS}, OVO^{WV}, OVO^{WWV}, and OVO^{NC} strategies, respectively, to enhance O-EBRBS. Meanwhile, Figs.

9(b)-9(d) show the changes in the number of rules, activated rules, and failed data for O-EBRBS, DRA-EBRBS, and Micro-EBRBS. Continuing with the case of O-EBRBS, the number of rules and activated rules is increased from 602.25 to 2807.96 and 338.36 to 1635.71, respectively, and the number of failed data is decreased from 2.50 to 0. On the basis of these improvement results, we can observe that the OVO strategy is able to not only enhance the capability of constructing a rule base, but also activate much more rules to produce a final output class. Additionally, the aggregation of the output of all binary EBRBSs is another notable merit of the OVO strategy, which contributes to avoid the situation of producing the failed data in the EBRB inference scheme, *i.e.*, for the classification problem with N classes, $N * (N - 1) / 2$ outputs can be produced because of $N * (N - 1) / 2$ binary EBRBSs and any one of these outputs is just one of basis to produce final output classes. Relatively, the EBRBS without the OVO strategy has to produce final output classes based on the only one output.



(a) Change in accuracies for three EBRBSs

(b) Change in number of rules for three EBRBSs

(c) Change in number of activated rules for three EBRBSs

(b) Change in number of failed data for three EBRBSs

Fig. 9. Influence of OVO strategy on O-EBRBS, DRA-EBRBS, and Micro-EBRBS

(2) Discussion of the influence of overlap functions on ERBBSs

The influence made by overlap functions is usually negative correction in the accuracy of the three EBRBSs. However, the implicit information can be found in the comparison and relationship among PR, SP, MI, GM, and SI functions. Taking O-EBRBS for example, Table 10 shows that the ranking order of the five overlap functions is 81.67% (SP) > 79.79% (PR) > 64.38% (MI) > 63.33% (GM) > 56.67% (SI). Obviously, this ranking order is the same to that of the five overlap functions to produce smaller $O(S_1^k,...,S_M^k)$, namely SP $\succ$ PR $\succ$ MI $\succ$ GM $\succ$ SI. Hence, we can observe that the overlap function, which is able to produce smaller $O(S_1^k,...,S_M^k)$, is more beneficial to improve the accuracy of EBRBSs. Based upon this viewpoint, the exponent of $O(S_1^k,...,S_M^k)$ is set as a = 1.0, 1.5, 2.0, 2.5, 3.0, 3.5, 4.0, 4.5, 5.0 to compare with the accuracy and its standard deviation of O-EBRBS, DRA-EBRBS, and Micro-EBRBS equipped by the PR function, as shown in Fig. 10, and the accuracy of O-EBRBS, DRA-EBRBS, and Micro-EBRBS equipped by PR, SP, MI, GM, and SI functions, as shown in Fig. 11.

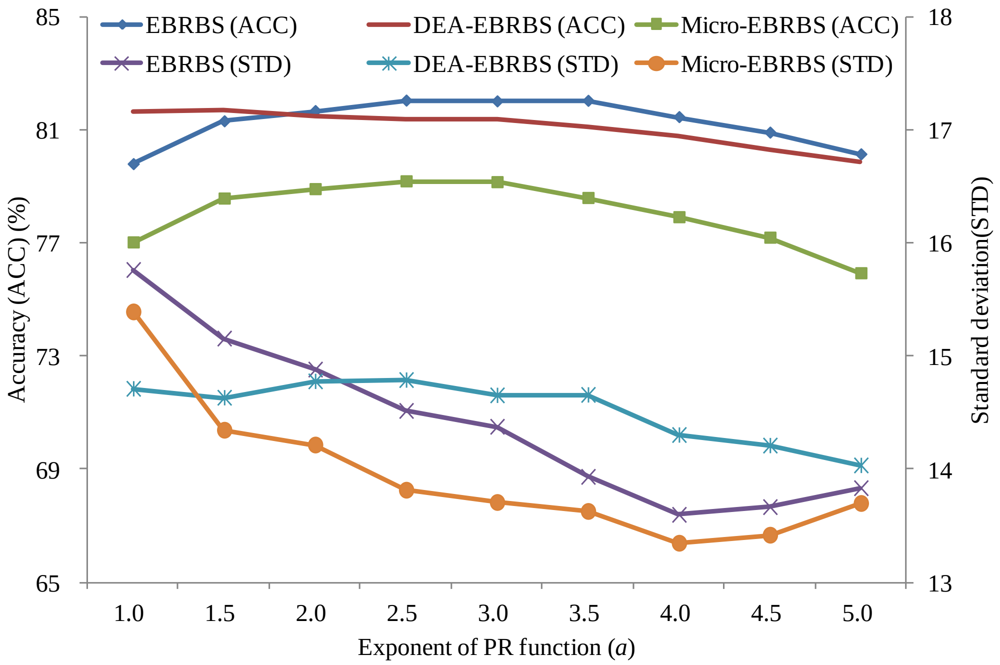
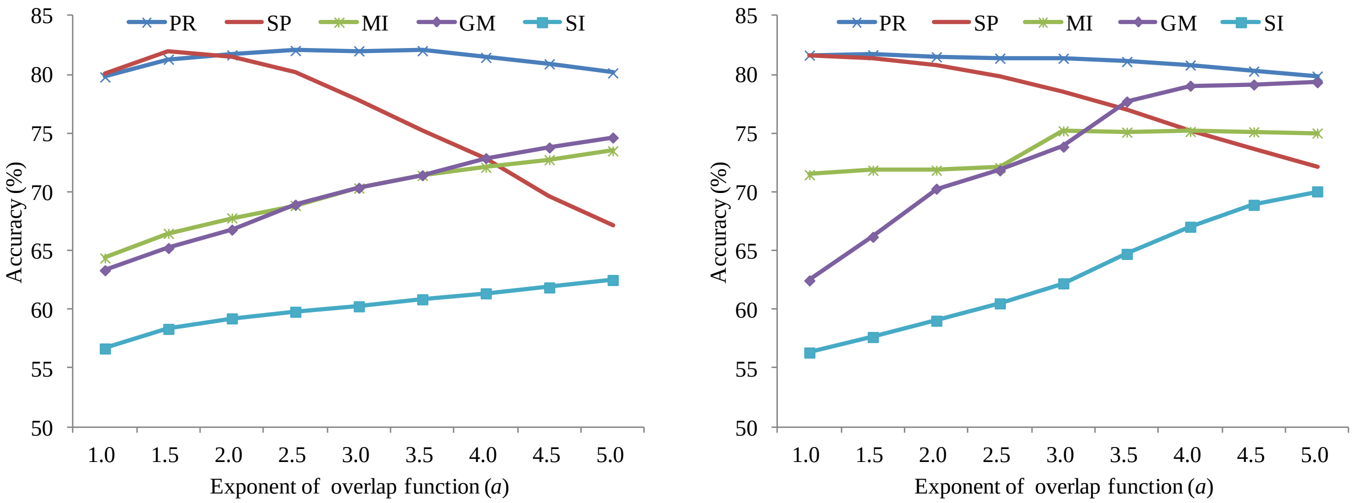


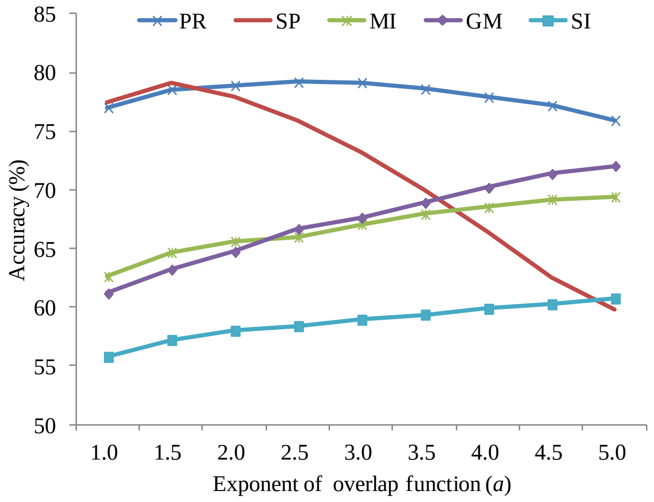
Fig. 10. Different exponents of PR function for the three EBRBSs

Fig. 10 shows that the accuracy of O-EBRBS and Micro-EBRBS initially increases and then decreases with the increase in the exponent of PR function. Meanwhile, the standard deviation of O-EBRBS and Micro-EBRBS decreases with the increase in the exponent of PR function. Thus, it can be concluded that there is a bound in the ability of overlap function to produce smaller $O(S_1^k,\ldots,S_M^k)$ and this bound is related with the best overlap function for improving the accuracy of O-EBRBS and Micro-EBRBS. The reason of having the bound is that activation weights tend to be zero with the increase of exponent, which means that O-EBRBS and Micro-EBRBS would fail to activate any rules for any given input data when there is a large exponent used for an overlap function. For DRA-EBRBS, whose accuracy and standard deviation decrease with the increase in the exponent of PR function, because the main idea of the dynamic rule activation method is to determine a best exponent of $O(S_1^k,\ldots,S_M^k)$ for consistent activated rules.



(a) Change in accuracies for O-EBRBS

(b) Change in accuracies for DRA-EBRBS



(c) Change in accuracies for Micro-EBRBS

Fig. 11. Different exponents of five overlap function for three EBRBSs

From Fig. 11, it can be found that the accuracy of O-EBRBS, DRA-EBRBS, and Micro-EBRBS with MI, GM, and SI functions increase with the increase of exponent. This is because MI, GM, and SI functions tend to produce a relatively large $O(S_1^k,...,S_M^k)$ comparing with SP and PR functions, which usually means that the former three functions are hard to determine consistent activated rules. In the same situation, the exponent of overlap functions can contribute to decrease the value of $O(S_1^k,...,S_M^k)$ and finally ensure that the activation weight of inconsistent rules would be zero. Comparatively, $O(S_1^k,...,S_M^k)$ of PR and SP functions is smaller than that of MI, GM, and SI functions, namely, their activation weights tend to be zeros with the increase of exponents. As a result, PR and SP functions usually fail to activate rules when having a large exponent. This is the reason why the accuracy of O-EBRBS, DRA-EBRBS, and Micro-EBRBS equipped by PR and SP functions initially increases and then decreases with the increase of exponent.

6. Conclusions

This work was motivated by the improvements found in some conventional classifiers when they worked with decomposition strategies and overlap functions for multi-attribute and multi-class classification problems. In this paper, the EBRBS, which has been proven to have a great potential in uncertainty information processing and modeling, was enhanced by using OVO strategy and five overlap functions. The work presented within this paper is meaningful and useful as a heuristic study to investigate the solution of improving the EBRBSs when handling multi-attribute and multi-class classification problems. The main conclusions can be summarized below:

(1) Three representative EBRBSs, namely O-EBRBS, DRA-EBRBS, and Micro-EBRBS, were investigated to establish their differences in term of accuracy and efficiency. The experimental study showed that DRA-EBRBS is an “accuracy first” classifier, Micro-EBRBS is an “efficiency first” classifier, and O-EBRBS is a “neutral” classifier.

(2) Five overlap functions, namely SP, PR, MI, GM, and SI functions, were investigated to determine which one is the most suitable for O-EBRBS, DRA-EBRBS, and Micro-EBRBS. The experimental study showed that SP function has a clear advantage in enhancing the classification performance of the three EBRBSs.

(3) Five OVO-related strategies, namely OVO^{VS}, OVO^{WV}, OVO^{WWV}, OVO^{NC}, and OVO^{LC} strategies, were investigated to determine which one is the most suitable for the three ERBBSs. The experimental study showed that OVO^{WV} strategy has a clear advantage in enhancing the performance of O-EBRBS and Micro-EBRBS.

There are several aspects that remain to be studied in future works, *e.g.*, the computing efficiency of decomposition strategies. Furthermore, an in-depth study of the effect of overlap functions on the interpretability of classifiers should be carried out.

Acknowledgments

This research was supported by the National Natural Science Foundation of China (Nos. 72001043 and 61773123), the Natural Science Foundation of Fujian Province of China (No. 2020J05122), the Humanities and Social Science Foundation of the Ministry of Education of China (No. 20YJC630188), the Social Science Planning Fund Project of Fujian Province of China (No. FJ2019C032), the Chengdu International Science Cooperation Project (No. 2020-GH02-00064-HZ), and the Spanish Ministry of Economy and Competitiveness through the Spanish National Research Project PGC2018-099402-B-I00.

References

[1] Jiang Z, Bian Z, Wang S, (2020). Multi-view local linear KNN classification: theoretical and experimental studies on image classification, *International Journal of Machine Learning and Cybernetics*, 11: 525-543. DOI: 10.1007/s13042-019-00992-9

[2] Yan W, Sun Q, Sun J, Li Y, (2020). Semi-supervised learning framework based on statistical analysis for image set classification, *Pattern Recognition*, 107: 1-15. DOI: 10.1016/j.patcog.2020.107500.

[3] Fan X, Hu S, He J, (2019). A dynamic selection ensemble method target recognition based on clustering and randomized reference classifier, *International Journal of Machine Learning and Cybernetics*, 10: 515-525. DOI: 10.1007/s13042-017-0732-2.

[4] Volna E, Kotyrba M, (2017). Enhanced ensemble-based classifier with boosting for pattern recognition, *Applied Mathematics and Computation*, 310: 1-14. DOI: 10.1016/j.amc.2017.04.019.

[5] Zhang DG, Wang JX, Fan HR, Zhang T, Gao JX, Peng Y, (2020). New Method of Traffic Flow Forecasting based on Quantum Particle Swarm Optimization Strategy for Intelligent Transportation System, International Journal of Communication Systems, 34(1): 1-20. DOI: 10.1002/dac.4647.

[6] Liu XH, Zhang DG, Yan HR, Cui YY, Chen L, (2019). A New Algorithm of the Best Path Selection based on Machine Learning. *IEEE Access*, 7: 126913-126928. DOI: 10.1109/ACCESS.2019.2939423.

[7] Galar M, Fernández A, Barrenechea E, Bustince H, Herrera F, (2011). An overview of ensemble methods for binary classifiers in multi-class problems: Experimental study on one-vs-one and one-vs-all schemes, *Pattern Recognition*, 44: 1761 - 1776. DOI: 10.1016/j.patcog.2011.01.017

[8] Liu Y, Bi JW, Fan ZP, (2017). A method for multi-class sentiment classification based on an improved one-vs-one (OVO) strategy and the support vector machine (SVM) algorithm, *Information Sciences*, 394-395: 38-52. DOI: 10.1016/j.ins.2017.02.016.

[9] Zhang T, Zhang DG, Yang HR, Qiu JN, Gao JX, (2021). A New Method of Data Missing Estimation with FNN-based Tensor Heterogeneous Ensemble Learning for Internet of Vehicle, *Neurocomputing*, 420: 98-110. DOI: 10.1016/j.neucom.2020.09.042.

[10] Zhang DG, Liu S, Liu XH, Zhang T, Cui YY, (2018). Novel Dynamic Source Routing Protocol (DSR) based on Genetic Algorithm- Bacterial Foraging Optimization (GA-BFO), *International Journal of Communication Systems*, 31(18): 1-20. DOI: 10.1002/dac.3824.

[11] Yang LH, Liu J, Wang YM, Martinez L, (2021). A Micro-Extended Belief Rule-Bases System for Big Data Multi-Class Classification Problems, *IEEE Transactions on Systems, Man, Cybernetics: Systems*, 51(1): 420-440. DOI: 10.1109/TSMC.2018.2872843.

[12] Zhang DG, Ge H, Zhang T, Cui YY, Liu XH, Mao GQ, (2019). New Multi-Hop Clustering Algorithm for Vehicular Ad Hoc Networks, *IEEE Transactions on Intelligent Transportation Systems*, 20(4): 1517-1530. DOI: 10.1109/TITS.2018.2853165.

[13] Gomez D, Rodriguez JT, Montero J, Bustince H, Barrenechea E, (2016). n-Dimensional overlap functions, *Fuzzy Sets and Systems*, 287: 57-75. DOI: 10.1016/j.fss.2014.11.023.

[14] De Miguel L, Gómez D, Rodríguez JT, Montero J, Bustince H, Dimuro GP, Sanz JA, (2019). General overlap functions, *Fuzzy Sets and Systems*, 372: 81-96. DOI: 10.1016/j.fss.2018.08.003.

[15] Calzada A, Liu J, Wang H, Kashyap A, (2015). A New Dynamic Rule Activation Method for Extended Belief Rule-Based Systems, *IEEE Transactions on Knowledge and Data Engineering*, 27(4): 880-894. DOI: 10.1109/TKDE.2014.2356460.

[16] Zhu HZ, Xiao MQ, Zhao X, Tang XL, Yang LH, Kang WJ, Liu ZZ, (2020). A structure optimization method for extended belief-rule-based classification system, *Knowledge-Based Systems*, 203: 1-15. DOI: 10.1016/j.knosys.2020.106096.

[17] Fang WJ, Gong XT, Liu GG, Wu YJ, Fu YG, (2020). A Balance Adjusting Approach of Extended Belief-Rule-Based System for Imbalanced Classification Problem, *IEEE Access*, 8: 41201-41212. DOI: 10.1109/ACCESS.2020.2976708.

[18] Liu J, Martinez L, Calzada A, Wang H, (2013). A novel belief rule base representation, generation and its inference methodology, *Knowledge-Based Systems*, 53: 129-141. DOI: 10.1016/j.knosys.2013.08.019.

[19] Zhang DG, Gong CL, Jiang KW, Zhang XD, Zhang T, (2019). A kind of new method of intelligent trust engineering metrics (ITEM) for application of mobile ad hoc network, *Engineering Computations*, 37(5): 1617-1643. DOI: 10.1108/EC-12-2018-0579.

[20] Cui YY, Zhang DG, Zhang T, Chen L, Piao MJ, Zhu HL, (2020). Novel method of mobile edge computation offloading based on evolutionary game strategy for IoT devices, *AEU - International Journal of Electronics and Communications*, 118: 1-10. DOI: 10.1016/j.aeue.2020.153134.

[21] Zhang DG, Zhao PZ, Cui YY, Chen L, Zhang T, Wu H, (2019). A New Method of Mobile Ad Hoc Network Routing Based on Greed Forwarding Improvement Strategy, *IEEE Access*, 7: 158514-158524. DOI: 10.1109/ACCESS.2019.2950266.

[22] Zhang DG, Chen L, Zhang J, Chen J, Zhang T, Tang YM, Qiu JN, (2020). A Multi-Path Routing Protocol Based on Link Lifetime and Energy Consumption Prediction for Mobile Edge Computing, *IEEE Access*, 8: 69058-69071. DOI: 10.1109/ ACCESS.2020.2986078.

[23] Yang LH, Wang YM, Su Q, Fu YG, Chin KS, (2016). Multi-attribute search framework for optimizing extended belief rule-based systems, *Information Sciences*, 370-371: 159-183. DOI: 10.1016/j.ins.2016.07.067

[24] Yang LH, Wang YM, Fu YG, (2018). A consistency analysis-based rule activation method for extended belief-rule-based systems, *Information Sciences*, 445-446: 50-65. DOI: 10.1016/j.ins.2018.02.059

[25] Zhang A, Gao F, Yang M, Bi WH, (2020). A new rule reduction and training method for extended belief rule base based on DBSCAN algorithm, *International Journal of Approximate Reasoning*, 119: 20-39. DOI: 10.1016/j.ijar.2019.12.016.

[26] Yang LH, Wang YM, Lan YX, Chen L, Fu YG, (2017). A data envelopment analysis (DEA)-based method for rule reduction in extended belief-rule-based systems, *Knowledge-Based Systems*, 123: 174-187. DOI: 10.1016/j.knosys.2017.02.021

[27] Alcala-Fdez J, Fernández A, Luengo J, Derrac J, García S, Sanchez L, Herrera F, (2011). KEEL Data-Mining Software Tool-Data Set Repository, Integration of Algorithms and Experimental Analysis Framework, *Journal of Multiple-Valued Logic and Soft Computing*, 17: 255-287. DOI: 10.1016/j.jlap.2009.12.002.

[28] García S, Fernández A, Luengo J, Herrera F, (2010). Advanced nonparametric tests for multiple comparisons in the design of experiments in computational intelligence and data mining-Experimental analysis of power, *Information Sciences*, 180: 2044 - 2064. DOI: 10.1016/j.ins.2009.12.010.

[29] Fernández A, Calderon M, Barrenechea E, Bustince H, Herrera F, (2010) . Solving multi-class problems with linguistic fuzzy

rule based classification systems based on pairwise learning and preference relations, *Fuzzy Sets and Systems*, 161: 3064-3080. DOI: 10.1016/j.fss.2010.05.016.

[30] Elkano M, Galar M, Sanz J, Fernández A, Barrenechea E, Herrera F, Bustince H, (2015). Enhancing Multiclass Classification in FARC-HD Fuzzy Classifier, On the Synergy between n-Dimensional Overlap Functions and Decomposition Strategies, *IEEE Transactions on Fuzzy Systems*, 23(5): 1562-1580. DOI: 10.1109/TFUZZ.2014.2370677.

[31] Garcia, LPF, Saez JA, Luengo J, Lorena AC, de Carvalho ACPLF, Herrera F, (2015). Using the One-vs-One decomposition to improve the performance of class noise filters via an aggregation strategy in multi-class classification problems, *Knowledge-Based System*, 90: 153-164. DOI: 10.1016/j.knosys.2015.09.023.

[32] Zhang Z, Krawczyk B, Garcìa S, Rosales-Pérez A, Herrera F, (2016). Empowering one-vs-one decomposition with ensemble learning for multi-class imbalanced data. Knowledge-Based Systems, 106: 251-263. DOI: 10.1016/j.knosys.2016.05.048.

[33] Liu JY, Jia BB, (2020). Combining One-vs-One Decomposition and Instance-Based Learning for Multi-Class Classification, IEEE Access, In Press. DOI: 10.1109/ACCESS.2020.3034448.

[34] Asmus TC, Dimuro GP, Bedregal B, Sanz JA, Jr. SP, Bustince H, (2020). General Interval-valued overlap functions and interval-valued overlap indices, *Information Sciences*, 527: 27-50. DOI: 10.1016/j.ins.2020.03.091.

[35] Arqub OA, AL-Smadi M, Momani S, Hayat T, (2016). Numerical solutions of fuzzy differential equations using reproducing kernel Hilbert space method, *Soft Computing*, 20(8): 3283-3302. DOI: 10.1007/s00500-015-1707-4.

[36] Arqub OA, Al-Smadi M, Momani S, Hayat T, (2017). Application of reproducing kernel algorithm for solving second-order, two-point fuzzy boundary value problems, *Soft Computing*, 21(23): 7191-7206. DOI: 10.1007/s00500-016-2262-3.

[37] Arqub OA, (2017). Adaptation of reproducing kernel algorithm for solving fuzzy Fredholm – Volterra integrodifferential equations, *Neural Computing & Applications*, 28: 1591-1610. DOI: 10.1007/s00521-015-2110-x.

[38] Arqub OA, Al-Smadi M, (2020). Fuzzy conformable fractional differential equations: novel extended approach and new numerical solutions, *Soft Computing*, 24(16): 12501-12522. DOI: 10.1007/s00500-020-04687-0.

[39] Chen JQ, Mao GQ, Li CL, Zhang DG, (2020). A Topological Approach to Secure Message Dissemination in Vehicular Networks, *IEEE Transactions on Intelligent Transportation Systems*, 21(1): 135-148. DOI: 10.1109/TITS.2018.2889746.

[40] Zhang DG, Chen C, Cui YY, Zhang T, (2018). New Method of Energy Efficient Subcarrier Allocation Based on Evolutionary Game Theory, *Mobile Networks and Applications*, 1: 1-14. DOI: 10.1007/s11036-018-1123-y.

[41] Zhang DG, Liu XH, Cui YY, Chen L, Zhang T, (2019). A kind of novel RSAR protocol for mobile vehicular Ad hoc network, *CCF Transactions on Networking*, 2(2): 111-125. DOI: 10.1007/s42045-019-00019-5.

[42] Zhang DG, Wu H, Zhao PZ, Liu XH, Cui YY, Chen L, Zhang T, (2020). New approach of multi-path reliable transmission for marginal wireless sensor network, *Wireless Networks*, 26(2): 1503-1517. DOI: 10.1007/s11276-019-02216-y.

[43] Elkano M, Galar M, Sanz J, Bustince H, (2016). Fuzzy Rule-Based Classification Systems for multi-class problems using binary decomposition strategies: On the influence of n-dimensional overlap functions in the Fuzzy Reasoning Method, *Information Sciences*, 332: 94-114. DOI: 10.1016/j.ins.2015.11.006

[44] Yang JB, Liu J, Wang J, Sii HS, Wang HW, (2006). Belief rule-base inference methodology using the evidential reasoning approach - RIMER, *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans*, 36(2): 266-285. DOI: 10.1109/TSMCA.2005.851270

[45] Kahraman HT, Sagiroglu S, Colak I, (2013), The development of intuitive knowledge classifier and the modeling of domain dependent data, *Knowledge-Based Systems*, 37: 283-295. DOI: 10.1016/j.knosys.2012.08.009

[46] Ye FF, Yang LH, Wang YM, Chen L, (2020). An environmental pollution management method based on extended belief rule

base and data envelopment analysis under interval uncertainty, *Computers & Industrial Engineering*, 144: 1-15. DOI: 10. 1016/j.cie.2020.106454.

[47] Wang YM, Ye FF, Yang LH, (2020). Extended belief rule based system with joint learning for environmental governance cost prediction, *Ecological Indicators*, 111: 1-14. DOI: 10.1016/j.ecolind.2020.106070

[48] Wang YM, Yang JB, Xu DL, (2006). Environmental impact assessment using the evidential reasoning approach, *European Journal of Operational Research*, 174(3): 1885-1913. DOI: 10.1016/j.ejor.2004.09.059

[49] Friedman J. Another approach to polychotomous classification Technical Report, Department of Statistics, Stanford University, 1996.

[50] Hüllermeier E, Vanderlooy S, (2010). Combining predictions in pairwise classification: an optimal adaptive voting strategy and its relation to weighted voting, *Pattern Recognition*, 43(1): 128-142. DOI: 10.1016/j.patcog.2009.06.013

[51] Hüllermeier E, Brinker K, (2008). Learning valued preference structures for solving classification problems, *Fuzzy Sets and Systems*, 159(18): 2337–2352. DOI: 10.1016/j.fss.2008.01.021

[52] Moreno-Torres JG, Saez JA, F. Herrera, (2012). Study on the Impact of Partition-Induced Dataset Shift on k-fold Cross-Validation, *IEEE Transactions on Neural Networks and Learning Systems*, 23(8): 1304-1312. DOI: 10.1109/TNNLS.2012. 2199516

[53] Zhang T, Zhang DG, Qiu JN, Zhang XD, Zhao PZ, Gong CL, (2019). A Kind of Novel Method of Power Allocation with Limited Cross-Tier Interference for CRN, *IEEE Access*, 7: 82571-82583. DOI: 10.1109/ACCESS.2019.2921310.

[54] Zhang DG, Gao JX, Liu XH, Zhang T, Zhao DX, (2019). Novel approach of distributed & adaptive trust metrics for MANET, *Wireless Networks*, 25(6): 3587-3603. DOI: 10.1007/s11276-019-01955-2.

[55] Zhang DG, Piao MJ, Zhang T, Chen C, Zhu HL, (2020). New algorithm of multi-strategy channel allocation for edge computing, *AEU - International Journal of Electronics and Communications*, 126: 1-10. DOI: 10.1016/j.aeue.2020.153372.

[56] Liu S, Zhang DG, Liu XH, Zhang T, Wu H, (2020). Adaptive repair algorithm for TORA routing protocol based on flood control strategy, *Computer Communications*, 151: 437-448. DOI: 10.1016/j.comcom.2020.01.024.