Applied Soft Computing Journal 91 (2020) 106214

Contents lists available at ScienceDirect

Applied Soft Computing Journal

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

A minimum centre distance rule activation method for extended belief rule-based classification systems

Haizhen Zhu [a,*], Mingqing Xiao [a], Longhao Yang [b], Xilang Tang [a], Yajun Liang [a], Jianfeng Li [a]

[a] ATS Lab, Air Force Engineering University, Xi'an, 710038, China
[b] Decision Sciences Institute, Fuzhou University, Fuzhou, 350116, China

ARTICLE INFO

Article history:
Received 15 May 2019
Received in revised form 30 December 2019
Accepted 2 March 2020
Available online 6 March 2020

Keywords:
Extended belief-rule-based system
Rule activation
Minimum centre distance
Classification
Filtering and selecting

ABSTRACT

Originating from the belief-rule-based (BRB) system, the extended belief rule-based (EBRB) system combined the advantages of the rule-based method and those of data-driven methods. By transforming the data set into extended belief rules and using evidential reasoning (ER), the EBRB system has expanded the application of BRB systems and demonstrated their capability in addressing classification problems. Nevertheless, the problem of activating nearly the entire rule base in every classification process is embedded in the EBRB scheme. There have been advances in rule activation for the EBRB system; however, the introduction of subjective information into the classification, high computational costs and long response times are common problems facing existing rule activation methods. To solve the problems facing rule activation for EBRB systems, a minimum centre distance rule activation (MCDRA) method for EBRB systems is proposed. In MCDRA, no subjective information is required, and no time-consuming iteration procedure is necessary. Two components of the proposed MCDRA, i.e., the filtering procedure and the selection procedure, are designed to eliminate unrelated samples of input query data and to select and activate the highly related samples to the input query data. A total of 12 benchmark data sets are used to test the performance of EBRB with MCDRA (M-EBRB). The experimental results show that compared with other rule activation methods, the proposed method obtains satisfactory rule activation ratios, accuracies and response times. Additionally, M-EBRB performs well on noisy data and comparatively with both the fuzzy-rule-based classification system (FRBCS) and several machine learning classification algorithms. In addition, MCDRA can be utilized as a generic rule activation method and can be used to optimize other rule-based classification systems.

1. Introduction

Classification is one of the most fundamental problems in the field of artificial intelligence and has various applications, including image identification [1], medical applications [2], fault diagnosis [3], and text categorization [4]. Despite its widespread application, classification still faces challenges in terms of higher classification accuracy and real-time performance. A number of machine learning algorithms, such as support vector machine [5] and k nearest neighbour [6], have demonstrated their effectiveness in addressing classification problems. Meanwhile, the rule-based method, which can utilize both expert knowledge [7,8] and information from data [9], has been rapidly developing, and attempts have been made to use the rule-based method to solve classification problems [10,11].

According to the aforementioned features of the rule-based method, this type of method can be generally divided into two basic categories: the methods whereby the rules are given by experts, in which human knowledge plays a critical role in the model, and data-driven methods, where the rule base is generated from sample data. For the former category, the models mainly focus on parameter learning to obtain the optimal performance [12–14]. The latter category, due to the large number of homogeneous rules, usually does not involve parameter learning and focuses on improving the computational efficiency [15,16]. However, to address classification problems, especially for high-dimensional and large-scale data sets, using parameter learning rule-based methods may be difficult [17]. Thus, data-driven rule-based methods with high computational efficiency are generally applied to classification problems.

Although there are a large number of applications with classification issues using data-driven rule-based methods, many methods also face challenges from combinatorial explosion problems because the numbers of rules increase exponentially with increasing numbers of attributes [18,19]. To settle this problem,

* Corresponding author.
E-mail addresses: haizhzhu@hotmail.com (H. Zhu), xmqing@sohu.com (M. Xiao), more026@hotmail.com (L. Yang), tangxilang@sina.com (X. Tang), 1214102891@qq.com (Y. Liang), 1506556759@qq.com (J. Li).

https://doi.org/10.1016/j.asoc.2020.106214
1568-4946/

several feature extraction methods, including grey target and principal component analysis, are analysed in [20] to decrease the number of attributes so that the scale of the rules can be reduced. However, in some cases, the interpretability of the rules may be affected when using feature extraction methods. Recently, Liu et al. proposed an extended belief-rule-based (EBRB) system in which each extended belief rule is generated from one sample of the data, therein solving the explosion problems without affecting the interpretability of the rules. In EBRB, the values of the sample data are transformed into belief distributions of attributes and the consequent part of the extended belief rules. The EBRB has demonstrated its potential in solving classification problems [21–23].

However, the problem of activating almost the entire extended belief rule base, which is embedded in the scheme of EBRB, would deteriorate the computing efficiency of EBRB [24]. Additionally, activating too many rules may introduce less input-related noisy data into the reasoning process of EBRB, influencing the classification accuracy of EBRB. Many studies have been performed to solve the rule activation problem of EBRB. In [25], an 80/20 method is proposed, where only the top 20% of the rules are activated. However, the 80/20 method may face big challenge for that the character of different data varies. Yang utilizes a k-dimensional tree-based search framework to determine the activated rules for small-scale attribute data sets [26], therein improving the efficiency of searching related rules. Nevertheless, the parameter k must be predefined subjectively. These methods can solve the problem of activating too many rules. However, introducing subjective information into the rule activation method introduces new drawbacks. In real world applications, the introduced parameters are subjectively determined by the experience of experts or users of these methods. The performance of these methods is sensitive to subjective information, which is unacceptable for a general method. In [27], a linear optimization consistency analysis-based rule activation (CABRA) model is developed to select appropriate activated rules without subjective information. However, the CABRA model is a time-consuming method because much time is needed to perform the linear optimization process. Thus, in the case study part of [27], the experimental results showed that CABRA improved the accuracy of EBRB. But the response time of CABRA is dramatic increased at the same time. A parameter tuning algorithm, named dynamic rule activation (DRA), is proposed in [28], where the interval of a parameter λ is given to control the number of activated rules. The λ is determined by the measure of incompleteness and inconsistency through a parameter adjusting process. Additionally, the range of λ would influence the performance of DRA. The wider range would be less restrictive in tuning λ and result in better accuracy. But, at mean time, the wider range of the parameter is, the longer computing time would be. The accuracy and response time of DRA have to be balanced when dealing with real-world problems. Additionally, in [29], Yang proposes a new activation rule determination method based on a linear separation of the activation region, where a linear programming model is embedded.

Summarizing the aforementioned rule activation methods, the common problem is either subjective information or time-consuming optimization process in the rule activation process. Motivated by the drawbacks mentioned before, a minimum centre distance rule activation (MCDRA) method is proposed in this paper. To avoid using subjective information and make the rule activation process free from time-consuming programming or iteration procedures are exactly the starting points of this paper. In MCDRA, a filtering procedure is designed to construct a hypersphere that contains all the sample data for each class. When an input query data is acquired, only the sample data from the classes whose corresponding hyperspheres cover the input query data are kept in this procedure. The filtering procedure aims to eliminate some unnecessary computing procedures in the subsequent steps in MCDRA. Then, a selection procedure is designed to select highly input-related sample data and activate their corresponding extended belief rules. In this procedure, a new hypersphere whose centre is the input query data is constructed. The sample data covered by this hypersphere are selected as candidate sample data. The selection and activation, which is based on distance, can be adjusted by the radius of this hypersphere. To guarantee that the number of activated extended belief rules is not influenced by the distribution of the data set, a distance adjustment factor based on distance density is introduced to adjust the radius of this hypersphere. Note that, this distance adjusting factor is determined by the distribution of the sample data and can be acquired by one computing step rather than using iteration algorithms. After adjustment, sample data within the coverage of the hypersphere are selected, and their corresponding extended belief rules are activated for integration using evidential reasoning (ER). All the procedures and steps of MCDRA are free from subjective information because the parameters and rule activation are determined by the data set rather than the user. Moreover, due to several time-saving designs in MCDRA, the computational cost of the proposed method remains very low, which makes it applicable for time-critical classification tasks.

To verify the validity of MCDRA, a case study is carried out using 12 benchmark data sets. The performance of EBRB with MCDRA (M-EBRB) is measured by the activation ratio, classification accuracy and response time. Moreover, a comparison experiment between M-EBRB and some popular machine learning algorithms is also conducted to further demonstrate the effectiveness of the proposed method.

The remainder of this paper is organized as follows: Section 2 briefly reviews the basic theory of EBRB and demonstrates problems existing in present methods. Detailed procedures of M-EBRB are illustrated in Section 3. Section 4 gives case studies whereby several comparison experiments are carried out to demonstrate the effectiveness of M-EBRB. Section 5 concludes the paper and provides directions for future research.

2. Review of extended belief rule base and problem demonstration

As an extension of Yang's belief rule-based (BRB) system [30], the EBRB system expands the uncertainty description of rule antecedents and consequents. Moreover, the extended belief rule base is generated from numerical sample data [31], which bridges the expert system method and data-driven method and is verified to be of outstanding effectiveness in addressing incomplete and fuzzy information [26]. The proposed MCDRA method is designed for the rule activation of EBRB, after which the activated rule from EBRB can be integrated through the ER algorithm. The following subsections review the basics of the EBRB and ER schemes.

2.1. Basics of EBRB

Let $\{U_1, U_2, \ldots, U_M\}$ be the antecedent attributes and $\left\{\left(A_{i,1}, \alpha_{i,j}^k\right)\right\}$ be the reference values and belief degrees of the ith antecedent attribute. $\left\{\left(D_n, \beta_n^k\right)\right\}$ are the corresponding consequents and belief degrees of the consequents. The belief structure is described in (1).

$$\begin{aligned} &R_k: \text{IF } U_1 \text{ is } \left\{\left(A_{1,j}, \alpha_{1,j}^k\right); j = 1, \ldots, J_1\right\} \wedge \cdots \wedge U_M \text{ is} \\ &\quad \left\{\left(A_{M,j}, \alpha_{M,j}^k\right); j = 1, \ldots, J_M\right\} \\ &\text{THEN } D \text{ is } \left\{\left(D_n, \beta_n^k\right); n = 1, \ldots, N\right\}, \\ &\quad \text{with rule weight } \theta_k \text{ and attribute weights } \{\delta_1, \ldots, \delta_M\}. \end{aligned} \tag{1}$$

where $\theta_k (0 \leq \theta_k \leq 1)$ is the weight of R_k and $\delta_i (0 \leq \delta_i \leq 1)$ denotes the importance of the ith antecedent attribute of the kth rule. $\alpha_{i,j}^k$ and β_n^k subject to $\sum_{j=1}^{J_i} \alpha_{i,j}^k \leq 1$ and $\sum_{n=1}^{N} \beta_n^k \leq 1$, respectively.

Remark 1. Parameter learning is not within the scope of our work. Both θ_k and δ_i are set to 1, which represents that all the extended belief rules generated from sample data and the attributes in each sample data are supposed to be equally important.

Remark 2. If we suppose that (1) reference values $A_{i,j}\,(i = 1, \ldots, M; j = 1, \ldots, J_i)$ are modelled by a certain membership function (such as a trapezoidal membership function for the fuzzy unordered rule induction algorithm (FURIA) and a triangular membership function for Chi-FRBCS), (2) $\alpha_{i,j}^k = 1\,(j = 1, \ldots, J_i)$ and $\alpha_{i,h}^k = 1\,(j = 1, \ldots, J_i, h \neq j)$, (3) $\delta_i = 1\,(i = 1, \ldots, M)$, and (4) $\beta_p^k = 1\,(p = 1, \ldots, N)$ and $\beta_q^k = 0\,(q = 1, \ldots, N, q \neq p)$, the kth extended belief rule changes into a fuzzy rule of FRBCS, which can be written as

$$R_k : \text{IF } U_1 \text{ is } A_{1,j} \wedge \cdots \wedge U_M \text{ is } A_{M,j}, \text{ THEN } D \text{ is } D_n, \text{ with rule weight } \theta_k. \tag{2}$$

Comparing the fuzzy rule and extended belief rule, we can determine that the fuzzy rule is a special version of the extended belief rule. Additionally, the extended belief rule is able carry uncertainty and incompleteness information in the attribute and consequent. For example, U_1 is $\{(low, 0.1)\,;\,(medium, 0.2)\,;\,(high, 0.7)\}$ means that for the first feature, we are 10% sure that it belongs to low, 20% sure that it belongs to medium and 70% sure that it belongs to high. $\{(D_1, 0.1)\,;\,(D_2, 0.7)\}$ means that there is 20% ignorance for the consequent part.

2.2. Transforming numerical sample data into belief distribution

As mentioned before, the EBRB is a data-driven method, which means that all the extended belief rules are generated by the labelled sample data. We assume that the kth numerical sample data sample is $\boldsymbol{x} = \{x_1, \ldots, x_M\}$ and that the utility value of the reference value $A_{i,j}$ is $u\left(A_{i,j}\right)$. The sample data are linearly transformed from sample data into belief distributions, as described in (3).

$$\begin{aligned} &S\left(x_{k,i}\right) = \left\{\left(A_{i,j}, \alpha_{i,j}^k\right); j = 1, \ldots, J_i\right\}, \\ &\quad \text{where } \alpha_{i,j}^k = \frac{u\left(A_{i,j+1}\right) - x_{k,i}}{u\left(A_{i,j+1}\right) - u\left(A_{i,j}\right)} \text{ and} \\ &\alpha_{i,j+1}^k = 1 - \alpha_{i,j}^k, \text{ if } u\left(A_{i,j}\right) \leq x_{k,i} \leq u\left(A_{i,j+1}\right). \alpha_{i,t}^k = 0 \\ &\quad \text{for } t = 1, \ldots, J_i \text{ and} \\ &t \neq j, j+1. \end{aligned} \tag{3}$$

The utility values of each attribute are generated evenly between the upper and lower boundaries of each attribute. To illustrate, we take 3 samples from the Iris data set, {4.8, 3.1, 1.6, 0.2, Iris-setosa}, {5.4, 3.0, 4.5, 1.5, Iris-versicolor} and {4.9, 2.5, 4.5, 1.7, Iris-virginica}, where the first four numbers in each sample denote attributes of the sample and the last description is the label of the data. We suppose that there are 3 utility values for each attribute. For simplicity, only the first two attributes are included in the example. The utility values and extended belief rules are generated as follows:

$U_1 = \{u\,(low), u\,(medium), u\,(high)\} = \left\{u\left(A_{1,1}\right), u\left(A_{1,2}\right), u\left(A_{1,3}\right)\right\} = \{4.8, 5.1, 5.4\}$

$U_2 = \{u\,(low), u\,(medium), u\,(high)\} = \left\{u\left(A_{2,1}\right), u\left(A_{2,2}\right), u\left(A_{2,3}\right)\right\} = \{2.5, 2.8, 3.1\}$

$D = \{\text{Iris-setosa, Iris-versicolor, Iris-virginica}\} = \{D_1, D_2, D_3\}$

Then, we have the following rules:

R_1: If U_1 is $\left\{\left(A_{1,1}, 1\right), \left(A_{1,2}, 0\right), \left(A_{1,3}, 0\right)\right\} \wedge U_2$ is $\left\{\left(A_{2,1}, 0\right), \left(A_{2,2}, 0\right), \left(A_{2,2}, 1\right)\right\}$, then D is $\{(D_1, 1), (D_2, 0), (D_3, 0)\}$

R_2: If U_1 is $\left\{\left(A_{1,1}, 0\right), \left(A_{1,2}, 0\right), \left(A_{1,3}, 1\right)\right\} \wedge U_2$ is $\left\{\left(A_{2,1}, 0\right), \left(A_{2,2}, 0.33\right), \left(A_{2,3}, 0.67\right)\right\}$, then D is $\{(D_1, 0), (D_2, 1), (D_3, 0)\}$

R_3: If U_1 is $\left\{\left(A_{1,1}, 0.67\right), \left(A_{1,2}, 0.33\right), \left(A_{1,3}, 0\right)\right\} \wedge U_2$ is $\left\{\left(A_{2,1}, 1\right), \left(A_{2,2}, 0\right), \left(A_{2,3}, 0\right)\right\}$, then D is $\{(D_1, 0), (D_2, 0), (D_3, 1)\}$

As described in the aforementioned example, all the extended belief rules can be generated from the numerical sample data, which means that the number of extended belief rules is exactly the number of data samples. This process maps the sample data into IF-THEN rules with minimal time consumption.

2.3. EBRB inference using ER

In the ER scheme, the rules that are related to the input query data are activated. Thus, the matching degrees between the input query data and all the extended belief rules generated in the aforementioned process must be evaluated. The ith antecedent attribute individual matching degree between the input query data and the kth rule can be calculated by

$$S^k\left(x_i, U_i\right) = 1 - \sqrt{\frac{\sum_{j=1}^{Ji}\left(\alpha_{i,j} - \alpha_{i,j}^k\right)^2}{2}} \tag{4}$$

where $\alpha_{i,j}^k$ is obtained by Eq. (3) and denotes the belief degree of sample data attribute U_i belongs to reference value $A_{i,j}$.

The activation weight of the kth rule is

$$w_k = \frac{\theta_k \prod_{i=1}^{M}\left(S^k\left(x_i, U_i\right)\right)^{\overline{\delta}_i}}{\sum_{l=1}^{L}\left(\theta_l \prod_{i=1}^{M}\left(S^l\left(x_i, U_i\right)\right)^{\overline{\delta}_i}\right)} \text{ and } \overline{\delta}_i = \frac{\delta_i}{\max_{i=1,\ldots,M}\{\delta_i\}}, \tag{5}$$

where θ_k and δ_i are the rule weight and attribute weights, respectively.

The estimated consequent belief distribution integrated from the activated rules is obtained by Eq. (6).

$$\beta_n = \frac{\mu * \left[\prod_{k=1}^{L}\left(\omega_k \beta_n^k + 1 - \omega_k \sum_{i=1}^{N} \beta_i^k\right) - \prod_{k=1}^{L}\left(1 - \omega_k \sum_{i=1}^{N} \beta_i^k\right)\right]}{1 - \mu * \left[\prod_{k=1}^{L}\left(1 - \omega_k\right)\right]} \tag{6}$$

where n = 1,..., N and

$$\mu = \left[\sum_{j=1}^{N}\prod_{k=1}^{L}\left(\omega_k \beta_j^k + 1 - \omega_k \sum_{i=1}^{N} \beta_i^k\right) - (N-1)\prod_{k=1}^{L}\left(1 - \omega_k \sum_{i=1}^{N} \beta_i^k\right)\right]^{-1} \tag{7}$$

The version of the consequent estimation is $\{(D_1, \beta_1), (D_2, \beta_2), \ldots (D_N, \beta_N)\}$. For classification problems, the consequents with the greatest belief are supposed to be the predicted label of the input query data.

$$p(x) = D_n, n = arg \max_{i=1,2,\ldots,N} \{\beta_i\} \tag{8}$$

2.4. Rule activation problems embedded in EBRB and existing works

In the conventional EBRB scheme, the rules whose activation weight is greater than 0 are activated in the ER process. However, as illustrated in [27], it is only in extreme conditions that the activation weight equals 0, being quite rare in real-world applications. Thus, in the ER process, nearly all the extended belief rules are activated and utilized to predict the label of the input query data. Many studies have attempted to address this problem. However, there are problems in the characteristics of the EBRB system and existing works.

Problem 1. The influence of noise and fewer input-related samples. The traditional EBRB scheme activates nearly all the extended belief rules, which enable it to fully utilize the information from the sample data. Nevertheless, in every sample data set, there is a substantial amount of noise data, which would deteriorate the performance of EBRB. Moreover, there are only a small fraction of rules that are highly related to the input query data. For example, in classification problems, the sample data are distributed as clusters. Only the clusters around the input query data are highly related to them. There is no need to activate the sample data from clusters far away.

Problem 2. The subjective information in some existing rule activation methods. Many works have attempted to address the rule activation problems of EBRB. However, there is subjective information embedded in the scheme determining the key parameters of existing rule activation methods. Activating the proper number of rules without subjective information is one of the topics of this paper. The detailed analysis of existing methods can be found in the introduction of this paper.

Problem 3. Time-consuming procedures that result in high computational costs for the classification process. In many real-world applications, there is a substantial amount of sample data collected for classification. Correspondingly, using conventional EBRB, the same number of rules would be generated from sample data. The activation of an excessive number of rules would reduce the efficiency of EBRB, making it unacceptable for real-time applications. For certain other rule activation methods, the activation procedure is determined after iterating parameter learning or programming procedures. These methods can select a tight subset of extended belief rules. Nevertheless, for some applications, the rule activation process is excessively time consuming. Thus, an efficient rule activation method is necessary for solving classification problems.

Summarizing the aforementioned problems, a method that can select a tight activation rule subset without the influence of subject information and a time-consuming process is necessary for EBRB. In the next section, a novel rule selection method that can overcome the existing problems mentioned above is proposed.

3. MCDRA method for EBRB

A MCDRA method is proposed in this section. There are two procedures in MCDRA: class filtering and highly input-related sample data selection. In the filtering procedure, the distances between the input query data and weighted centres are utilized to filter classes that are obviously unrelated to the input query data. Thus, we are able to avoid activating too many extended belief rules, and many unnecessary calculations can be skipped, which greatly reduces the computational cost. In the selection procedure, sample data from the filtered class are further selected as the activated sample data to generate the extended belief rules.

3.1. Filtering the unrelated classes

In the filtering procedure, we have sample data from a variety of classes. Suppose that there are p_i samples in the ith class and that the attributes and consequent of the sample data $S_{i,j},\ j = 1, 2, \ldots, p_i$ are $\left\{x^i_{j,1}, x^i_{j,2}, \ldots, x^i_{j,m}\right\}$ and D_i, respectively. The weighted centre of the ith class can be obtained by

$$\begin{aligned} C_i &= \left\{xc^i_{j,1}, xc^i_{j,2}, \ldots, xc^i_{j,m}\right\} \\ &= \left\{\sum_{j=1}^{p_i} x^i_{j,1}/p_i, \sum_{j=1}^{p_i} x^i_{j,2}/p_i, \ldots, \sum_{j=1}^{p_i} x^i_{j,m}/p_i\right\} \end{aligned} \tag{9}$$

The distance between points p and q is denoted by $d_{p,q}$. Thus, the Euclidean distance from the sample to the weighted centre can be calculated by

$$d1_{i,j} = \sqrt{\sum_{a=1}^{m} \left(x^i_{j,a} - xc^i_{j,a}\right)^2} \tag{10}$$

The Euclidean distance from the input query data to the weighted centre is

$$d2_{i,q} = \sqrt{\sum_{a=1}^{m} \left(xq^i_{j,a} - xc^i_{j,a}\right)^2} \tag{11}$$

If $\max d1_{i,j} \le d2_{i,q}$, the input query data are supposed to not belong to the ith class. Thus, the sample data from the ith class are abandoned when classifying these input query data. If $\max d1_{i,j} \ge d2_{i,q}$, the sample data from the ith class are temporarily retained for the selection procedure. In the selection procedure, we construct a hypersphere for each class whose centre is the weight centre. Its radius is the greatest distance from its centre to the sample belonging to this class. If the input query data is within the coverage of this hypersphere, the sample data from this class are retained.

The filtering procedure is illustrated in Algorithm 1:

Algorithm 1 Minimum Centre Distance Rule Activation Method: Filtering

Input: All the sample data $S_{i,j}, i = 1, 2, \ldots, p,\ j = 1, 2, \ldots, p_i$ with attributes $\left\{x^i_{j,1}, x^i_{j,2}, \ldots, x^i_{j,m}\right\}$ and consequents D_i; Input query data x_q;

Output: Filtered sample data set A;

```
A ← Φ
for i = 1 → p do
    Calculate weighted centres of sample data from the ith class C_i;
    for j = 1 → p_i do
        Calculate the distance between sample data and their corresponding weighted centres d1_{i,j};
    end for
    L1_i ← max d1_{i,j};
end for
Calculate the distances between input query data and the ith weighted centre d2_{i,q};
if L1_i ≥ d2_{i,q} then
    A ← A ∪ S_{i,j};
end if
return A;
```

Remark 3. As we can see in Algorithm 1, the first 8 steps can be executed one time for one data set after the sample data are acquired. Specifically, the weighted centre and the distances from the sample data to the weighted centre can be calculated ahead of obtainment of the input query data. Moreover, the distance calculation is less time consuming. Thus, the computational cost is very low for the filtering procedure.

Note that the query-unrelated sample data are a relative concept. In the classification process, sample data from a discarded class in the filtering process for one query data sample may be kept and activated for other data after the filtering and selecting procedures are performed.

In the filter process of our method, the first several steps, which are generating hypersphere for each class, are exactly the same as that in supervised evidential C-means clustering (ECM) [32]. The purpose of rule activation for EBRB is to select the input-related extended belief rules. While for CBRBCS, the aim of

supervised ECM is go get the best data partition so as to generate the belief rules. As a result, the following steps are different. In filter, the hypersphere (the label-determined initial cluster in [32]) is constructed to obtain the covering area of each class. Then in filter is designed to eliminate the input-unrelated classes and the rest classes are kept for further process by the selection algorithm. For the supervised ECM, the following steps are parameter optimization to get the final partition which are different from the partition determined by the label. Additionally, one of the starting point of our paper is to downsize the computing cost of rule activation process for EBRB. As a preprocessing procedure of the following selection algorithm, the filter is designed to efficiently deal with data. The rule activation is mainly accurately conducted by selection algorithm. Therefore, the optimization process is not involved in filter, which is different from that of supervised ECM.

3.2. Selecting highly related sample data in the filtered class

In the selection procedure, a tighter subset of filtered sample data is to be determined. There are 3 potential cases after the filtering procedure.

Case 1: Sample data are not retained after filtering. Under this situation, the input query data is not within the coverage of the hyperspheres of all classes. In this case, sample data from class i with minimum $d2_{i,q}$ are activated. Note that the labels of all the activated sample data are the same. Using ER, the μ in Eq. (7) is equal to 1. We substitute $\mu = 1$ into Eq. (6); the belief distribution of β_n is

$$\beta_n = \begin{cases} 0 & n = 1, 2, \ldots, N, n \neq i \\ 1 & n = i \end{cases} \tag{12}$$

where i denotes the retained class. Thus, the predicted label of the input query data, which is denoted as D_q, is supposed to be D_i.

Case 2: The retained sample data belong to the same class, i.e., the input query data is within the coverage of the hypersphere of one class. When using ER, this case is exactly the same as case 1. Thus, D_q is supposed to be D_i, where i is the retained class.

Case 3: The retained sample data are from more than one class. The input query data is in the meta-class section (overlapping coverage area of several classes) [33]. If all the retained sample data are selected, it is very likely that the classification is largely determined by the amount of sample data with less influence by the distance. Some of the sample data that are far from the input query data are retained in the filtering procedure due to the maximum sample-centre scheme of the filtering procedure. These sample data are usually on the edge of the retained class, and some data are in the opposite direction of the weighted centre and the input query data. These sample data would deteriorate the performance of ER in classifying the input query data. To eliminate the influence of these sample data, the selection scheme of Case 3 is as follows:

Step 1: Calculate the Euclidean distance between the retained sample data and input query data $d3_{k,q}$.

We suppose that there are sample data from M classes that are retained after the filtering procedure. The weighted centres of the retained class are $w_1, \ldots, w_M$. The minimum centre distance of the input query data is $\min d_{q,w_i}$, $i = 1, \ldots, M$.

Step 2: If $d3_{k,q}$ is less than $\min d_{q,w_i}$, the sample data S_k are retained as a candidate sample.

Note that the number of candidate samples retained in step 2 is influenced by the local distribution of sample data that are around the input query data. For example, if an input query data is located at a point where the local sample data are quite densely distributed, using the same $\min d2_{i,q}$, the amount of candidate sample data retained in step 2 would be much greater than that of input query data, which are located at sparsely distributed local sample data points. The local outlier factor (LOF) is a factor that can monitor the degree of being an outlier for each sample data [34]. One method for calculating the LOF for a sample is the distance density. Readers can refer to [35] for detailed information. To compensate for the local distribution difference of the sample data, a minimum distance density adjustment factor is introduced in our method to select activated samples in candidate samples.

The minimum distance density adjustment factor for input query data can be calculated through the k-distance and distance density.

Definition 1. The k-distance ($D_k(p) = d_{p,o}$): the distance of point p to its kth nearest neighbour, excluding p.

Remark 4. We call all points within the k-distance of p, including the k-distance, the k-distance neighbourhood of p, which is denoted by $N_k(p)$.

Definition 2. Distance density: the reciprocal of the average distance from points that are covered by the k-distance neighbourhood of point p to point p. The distance density of point p can be calculated by

$$den(p) = \frac{1}{\sum_{o \in N_k(p)} d_{p,o} / \mid N_k(p) \mid} \tag{13}$$

where $\mid N_k(p) \mid$ is the number of samples in the k-distance neighbourhood of p and $\mid N_k(p) \mid = k$.

Remark 5. The distance density is used to represent the local point density. The greater the distance density of a point is, the denser the distribution of points around this point. The computational cost of calculating the distance density linearly increases with increasing dimension.

In MCDRA, the minimum centre distance $\min d_{q,w_i}$ of input query data is regarded as the k-distance of the input query data q. In other words, the number of data samples covered by the hypersphere whose centre and radius are the input query data and the k-distance of the input query data q, respectively. The number of k in the k-distance for each input query data sample is determined by the data.

Let CN_i be the number of samples in class i. The average class distance density after the filtering procedure is

$$Cden_{ave} = \frac{1}{(\sum_{i=1}^{M} \sum_{o \in class\ i} d_{w_i,o}) / (\sum_{i=1}^{M} CN_i)} \tag{14}$$

The minimum distance adjustment factor is

$$f = \frac{Cden_{ave}}{den(q)} \tag{15}$$

The minimum distance adjustment factor is used to adjust the radius of the new hypersphere. Thus, the final step of Case 3 is as follows:

Step 3: If $d3_{k,q}$ is less than $f * \min d_{q,w_i}$, where $i = 1, \ldots, M$, the sample data S_k are selected to generate the extended belief rules. These rules are supposed to be the activated rules for the ER procedure.

Summarizing the aforementioned procedures, the pseudocode of the selection procedure is as shown in Algorithm 2:

Algorithm 2 Minimum Centre Distance Rule Activation Method: Selection

Input: Input query data x_q; The distances between the input query data and the ith weighted centre $d2_{i,q}$; Filtered sample data set, A;

Output: Predicted consequent input query data D_q or activated sample set R;

if $A == \Phi$ **then**
  $D_q \leftarrow$ class i with min $d2_{i,q}$;
else if {the number of class in A} == 1 **then**
  $D_q \leftarrow$ the class i in A ;
else
  $K \leftarrow sizeof(A)$
  Calculate the distance between input query data and sample data in A: $d3_{k,q}, k = 1, 2, \ldots, K$;
  while *classi* $\in A$ **do**
    Calculate the distance density of the weighted centre of each class and input query data;
    Calculate the minimum distance adjustment factor f for q
    for $k = 1 \rightarrow K$ **do**
      if $f * min\ d2_{i,q} \geq d3_{k,q}$ **then**
        $R = R \cup S_k$
      end if
    end for
  end while
end if
return D_q or R;

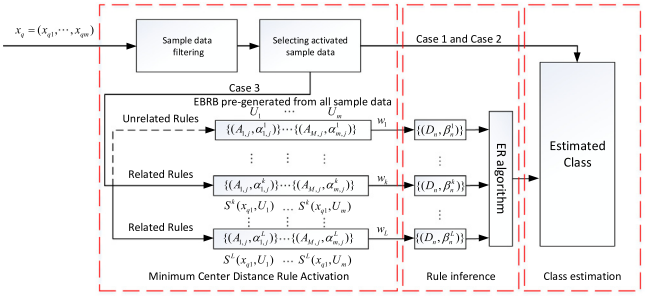

Fig. 1. Integrating MCDRA with EBRB.

3.3. Integrating the MCDRA with EBRB

The MCDRA model, including the filtering and selection framework, is designed to activate appropriate extended belief rules for the ER algorithm to estimate the class of the input query data. The scheme of M-EBRB is described in Fig. 1.

Step 1. Activation rule selection. To decrease the computing time, the extended belief rules are generated from sample data by Eq. (3) before MCDRA. The weighted centres of sample data from different classes can be calculated through (9).

Step 2. In the filtering process, the classes that are regarded as being relative to the input query data are selected. After that, in the selection process, a tighter subset of the filtered sample data is selected. If no sample data are selected or only sample data from one class are selected (case 1 and case 2), the estimated class is given in this step; otherwise, the class is regarded as a meta-class. The detailed procedures are illustrated in Algorithms 1 and 2. The corresponding rules of the selected sample data are activated.

Step 3. Calculating the activation weights for the selected rules. In our work, the similarity calculation method proposed by Yang is adopted to avoid counterintuitive problems. For a detailed description of the counterintuitive problem, please see [36]. The new similarity between the ith attribute of the input sample data and the ith attribute of the kth rule is $\overline{S^k(x_i, U_i)}$, as depicted in (17). Eqs. (16) and (17) are used to calculate the individual matching degree between the input query data and activated rules. Then, the activation weights of the extended belief rules can be obtained through Eq. (5) by replacing $S^k(x_i, U_i)$ with $\overline{S^k(x_i, U_i)}$ in [36].

$$\gamma_{i,j} = 1.0 - \frac{\left|u\left(A_{i,j}\right) - x_i\right|}{\max_{k=1,\ldots,J_i}\left\{u\left(A_{i,k}\right)\right\} - \min_{k=1,\ldots,J_i}\left\{u\left(A_{i,k}\right)\right\}}, \tag{16}$$

where $u\left(A_{i,k}\right)$ denotes the utility value of $A_{i,j}$.

$$\overline{S^k(x_i, U_i)} = \sum_{j=1}^{Ji} \gamma_{i,j}\alpha_{i,j}^k \tag{17}$$

where $\alpha_{i,j}^k$ is obtained by Eq. (3).

Step 4. Class estimation of input query data. All the activated rules are integrated to estimate the inference output using ER. Note that in this step, L refers to the number of activated rules. The label of the input query data is estimated to be the class (n) with the largest β_n.

Remark 6. In [37] and [32], a compact belief rule-based classification system (CBRBCS) is proposed, where the initial partition and number of clusters is determined by the label of data. By introducing the ECM to the framework of BRB proposed in [16], the partition of data is obtained and the data in the same partition are fused to generate the rules. The framework of BRB in [16] (as well as CBRBCS) differs from that of the EBRB proposed in [31]. The ECM is used to give the initial partition and obtain the final partition through optimization. The framework of our selection algorithm is based on the EBRB framework, which is proposed in [31], and improved in [28] (DRA) and [29]. In our framework, the extended belief rules are generated directly from original data, where each rule has its corresponding data. Afterwards, a proper number of extended belief rules are integrated in the evidential reasoning process. The starting point of our work, as we have stated in Section 1 is to address the problems existing in the rule activation methods. So the selection algorithm of MCDRA is designed to efficiently select and activate input-related extended rules which differs from that of ECM. The EBRB system has the property of fast generation of extended belief rules. The selection algorithm of MCDRA, which activate proper input-related and smaller number of extended belief rules, further contribute to the performance of this framework in terms of response time and accuracy.

4. Case studies

To verify the effectiveness of the MCDRA method, 12 benchmark data sets are downloaded from the UCI repository of machine learning databases [38]. First, the Iris data set is used as an example to illustrate the working scheme of MCDRA. Next, 4 experiments are conducted to illustrate the effectiveness of M-EBRB. In the first experiment, to verify the rule activation performance of M-EBRB, the rule activation ratios of conventional EBRB (abbreviated as C-EBRB) and M-EBRB of 12 data sets are given under different conditions. In the second experiment, the C-EBRB, DRA-based EBRB (abbreviated as D-EBRB) [28] and the activation method proposed in [29] (abbreviated as N-EBRB) are used for comparison. Afterwards, the performance of M-EBRB under different levels of attribute noise and label noise as well as that of FRBCS are given. Finally, a comparison of the classification

Table 1
Description of data sets.

Data sets	No. of instances	No. of attributes	No. of classes
Iris	150	4	3
Seeds	210	7	3
Glass	214	9	6
Ecoli	336	7	8
Diabetes	393	8	2
Cancer	569	30	2
Mammographic	830	5	2
Banknote	1372	4	2
Yeast	1484	8	10
Titanic	2201	3	2
Banana	5300	2	2
Phoneme	5404	5	2

accuracy between M-EBRB and state-of-the-art machine learning algorithms is provided.

4.1. Data sets and experimental settings

Detailed information of the data sets is given in Table 1. Note that samples in the data sets with missing attribute values are removed.

When using EBRB methods, the utility values of each attribute are generated evenly between the upper and lower boundaries of each attribute. N-fold cross-validation (N-CV) is utilized, which denotes that each data set is randomly divided into N parts, and we take $N-1$ parts to generate an extended rule base and 1 to act as a test set. The experiment was conducted in MATLAB R2014b on an Intel(R) Core(TM) i7-6500 U CPU 2.50 GHz using Windows 10.

4.2. Illustration of M-EBRB scheme

To illustrate the rule activation and ER scheme of M-EBRB, the Iris data set is taken as an example. As shown in Table 1, each sample has four attributes: sepallength, sepalwidth, petallength and petalwidth. There are 3 classes (Iris-setosa, Iris-versicolor and Iris-virginica) in the Iris data set. In this illustration, 150 samples are randomly divided into 10 parts, where 9 parts (135 samples) are used to generate EBRB, and the samples in the other part are used as input query data. To make the illustration more intuitive, we only take the last two attributes. We will illustrate the most common situation, Case 3, of M-EBRB in this subsection.

Step 1: According to the scheme of M-EBRB, first, EBRB is generated from 135 samples. After that, 3 weighted centres of the original sample data can be obtained. As shown in fig (distribution), the red crosses, the green plus signs and the magenta asterisk represent samples from Iris-setosa, Iris-versicolor and Iris-virginica, respectively. The 3 black diamonds are weighted centres for each class, and the blue pentagram represents input query data in Fig. 2.

Step 2: In the filtering procedure of M-EBRB, we calculated the distance of each data sample to its corresponding weighted centre. The longest distance is taken for the radius of each class. In this example, the radii for Iris-setosa, Iris-versicolor and Iris-virginica are 0.46, 1.24 and 1.35, respectively. Then, we calculate the distance from the input query data to the three weighted centres. Then, the samples from the classes whose circle (a hypersphere in a high-dimensional space) covers the input query data would be retained in the filtering procedure. As can be observed in Fig. 3, samples from Iris-versicolor and Iris-virginica are retained when classifying these input query data.

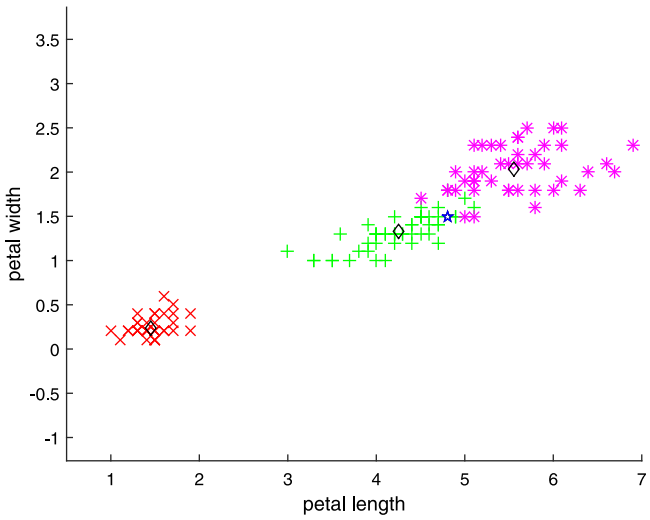

Fig. 2. Distribution and weighted centres of sample data.

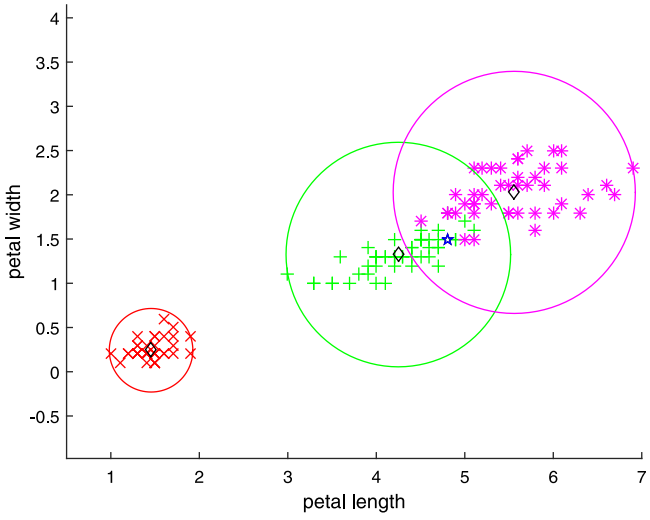

Fig. 3. Filtering procedure of MCDRA.

In the first step of selection procedure, the candidate samples are determined by the minimum distance between the input query data and the weighted centres of the retained class in the filtering procedure (distance between the input query data and weighted centre of Iris-versicolor in this example). The corresponding minimum distance in the example is 1.24, which is set as the radius of a new circle whose centre is the input query data. The points covered by this circle (the black dashed line) are the candidate sample data. There are 26 sample data (17 from Iris-versicolor and 9 from Iris-virginica) within the coverage of this circle. Thus, the k for the kth neighbourhood is 26 in this example (see Fig. 4).

The average distance of the Iris-versicolor and Iris-virginica is 1.3018, and the distance density of the input query data is 1.5152. Thus, the minimum distance adjustment factor is 0.8620, and the radius of the final circle is $0.8620 * 1.24 = 1.0689$. As shown in Fig. 5, the blue dashed circle is the final circle, and the corresponding extended rules of the 23 sample data from Iris-versicolor(15) and Iris-virginica(8), which are covered by the final circle, are activated for ER. The blue dashed circle is smaller than the black circle. This is because the example query data (marked with a star) is located in a densely distributed area of relative

Table 2
Rule activation ratio (%) of C-EBRB and M-EBRB.

Data set	No. of sample	2-CV		5-CV		10-CV	
		C-EBRB	M-EBRB	C-EBRB	M-EBRB	C-EBRB	M-EBRB
Iris	150	94.13	26.05	92.08	24.71	93.93	23.77
Seeds	210	99.81	22.86	99.94	20.56	99.89	19.71
Glass	214	99.16	3.82	98.77	3.20	92.40	3.33
Ecoli	336	97.62	20.34	97.09	19.82	96.42	19.26
Diabetes	393	99.84	8.41	99.15	7.84	99.91	7.62
Cancer	569	100	10.44	100	9.78	100	9.49
Mammographic	830	84.52	5.52	74.60	5.44	72.71	5.20
Banknote	1372	100	6.53	100	6.20	100	6.33
Yeast	1484	98.19	6.10	98.10	5.89	98.07	5.82
Titanic	2201	100	22.43	100	22.62	100	22.45
Banana	5300	100	1.71	100	1.70	100	1.68
Phoneme	5404	100	0.52	100	0.44	100	0.45
Average		97.77	11.23	96.64	10.68	96.11	10.43

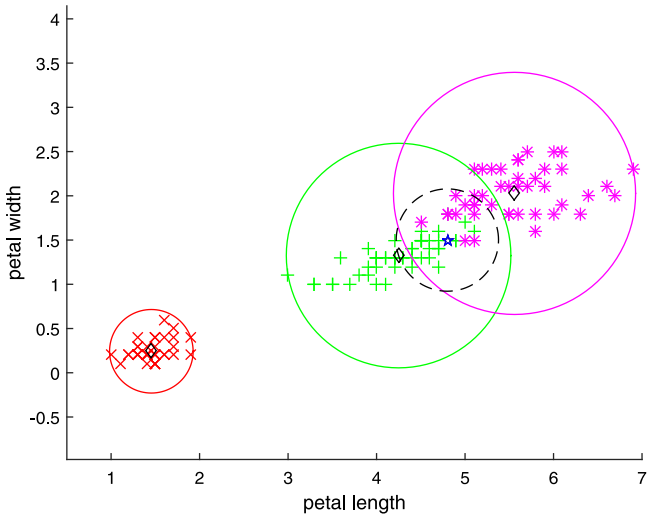

Fig. 4. The candidate sample data.

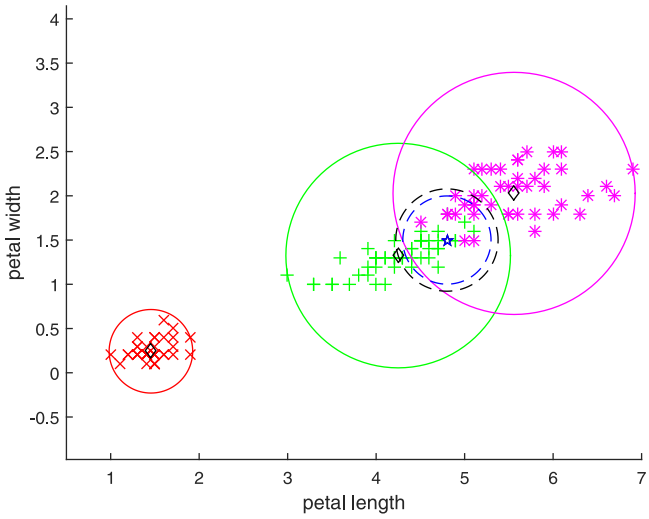

Fig. 5. Selection procedure of MCDRA.

sample data whose sample data density is slightly higher than the average for all data. Thus, under the influence of the density adjustment factor, the selected area is narrowed.

Step 3: Calculate $\overline{S^k}(x_i, U_i)$ using Eq. (17) and w_k using Eq. (5) for 23 activated rules.

Step 4: Integrate 23 activation weights and their corresponding consequent belief distribution, where the predicted label of the input query data is $(D_1, 0), (D_2, 0.79), (D_3, 0.21)$. The input query data is classified as Iris-versicolor.

4.3. Rule activation ratio comparison

To illustrate the effectiveness of M-EBRB for rule activation, we conduct comparative experiments between C-EBRB and M-EBRB concerning the rule activation ratio under the conditions of 2-CV, 5-CV and 10-CV. Under different CV settings, we randomly divided the original data sets into corresponding numbers of parts. For example, when conducting an experiment on Iris using 5-CV, we divided the data set into 5 parts (30 sample data in each part). Then, in turn, we selected 4 parts to generate the extended belief rules and used data from the remainder as input query data. Thus, by calculating the number of activated extended belief rules and averaging the outcomes of each experiment, we can obtain the rule activation ratio of C-EBRB and M-EBRB. Except for the random division in every experiment, the results of the 12 data sets are averaged from 10 experiments under the same condition. The results are shown in Table 2.

The experiments are conducted under the same input settings. The common feature is that the greater the number of CV folds is, the slightly lower the activation ratio. This may result from the occurrence ratio of extreme conditions, where the individual matching degree equals 0, increasing more so than the increase ratio of rule-generating samples. Regarding the differences, the activation ratio of M-EBRB is greatly reduced compared to C-EBRB. This is because fewer input-related samples are eliminated during the filtering and selection process, which would contribute to the lower response time. For C-EBRB, it is quite clear that it would activate most of the extended belief rules. The average rule activation ratio for the 12 data sets is 97.77%, 96.64% and 96.11% under 2-, 5-, and 10-CV, respectively. Additionally, note that for Mammographic, the activation ratio of C-EBRB ranges from 72.71% to 84.52%, which is quite low compared to the other data sets. This is because the attribute values of Mammographic are several fixed discrete values for each attribute, which would result in that part of the individual matching degrees equalling 0. However, this situation only occurs in special data sets. For the other data sets, the individual matching degrees are greater than 0 except for extreme conditions. It can be observed that for Cancer, Banknote, Titanic, Banana and Phoneme, the activation ratio is 100%. In other words, all the extended belief rules are

Table 3
Accuracy (%) of EBRB methods.

Data set	No. of sample	C-EBRB	D-EBRB	N-EBRB	M-EBRB
Iris	150	95.2	95.5	95.73	97.80
Seeds	210	87.04	92.02	93.24	93.45
Glass	214	51.43	69.65	75.51	74.36
Ecoli	336	33.72	83.76	87.14	87.24
Diabetes	393	73.39	71.44	75.98	75.17
Cancer	569	94.59	94.61	95.47	95.79
Mammographic	830	77.64	78.39	79.57	80.14
Banknote	1372	98.53	99.70	98.83	99.65
Yeast	1484	30.91	54.13	54.51	58.62
Titanic	2201	76.74	78.83	78.06	78.91
Banana	5300	63.43	87.06	87.28	89.82
Phoneme	5404	71.89	76.02	85.23	87.17
Average		71.21	81.76	83.88	84.84

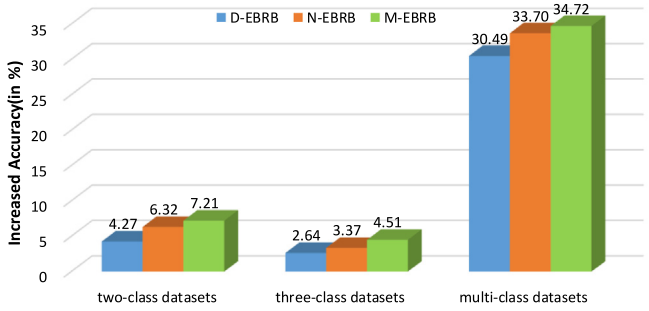

Fig. 6. Increased accuracy of different rule activation methods compared to C-EBRB.

to be utilized in the ER procedure, which contributes to the lack of efficiency of C-EBRB. In comparison, for all 12 data sets and all 3 cross-validation settings, M-EBRB maintains a relatively low rule activation ratio. On average, the activation ratio is nearly 10%, which largely decreases the computational cost of the algorithm.

4.4. Accuracy and response time comparison among EBRB methods

In this experiment, 10-CV is adopted, and the experimental results of M-EBRB are averaged from 20 duplicated experiments for the 12 data sets. First, the performance of the EBRB methods is measured in terms of the accuracy, which is the ratio of samples that are correctly classified from test data. The other performance metric is the response time (in milliseconds), which is the time required to classify each input query data.

The experimental results in terms of accuracy for the 12 data sets are shown in Table 3. The best performances are underlined.

In terms of accuracy, D-EBRB, N-EBRB and M-EBRB outperform the C-EBRB for most of the data sets, which indicates that input-related extended belief rule selection methods not only ease the computational cost of the ER procedure but also increase the classification accuracy.

For small-scale data sets (with fewer than 1000 samples), the N-EBRB obtains the best accuracy on 2 data sets, and the number of that of M-EBRB is 5. By analysing the accuracy further, we can infer that M-EBRB has comparable performance to N-EBRB. For large-scale data sets, the accuracy of M-EBRB in classifying Banknote is 99.65%, which is slightly lower than when using D-EBRB (99.70%). Nevertheless, for the other large-scale data sets, M-EBRB achieves the best performance of all four methods. In general, the proposed M-EBRB, whose average is 84.84%, performs better than D-EBRB and N-EBRB.

Table 4
Wilcoxon's test of accuracy for comparison of M-EBEB with other rule activation methods ($\alpha = 0.05$).

Methods	z-value	p-value	Hypothesis
C-EBRB	−3.059	2.218 E-3	Rejected
D-EBRB	−2.981	2.873 E-3	Rejected
N-EBRB	−2.040	4.139 E-3	Rejected

Compared with the results of C-EBRB, M-EBRB performs better on all 12 data sets. For Glass, Ecoli and Yeast, the accuracy increased by 22.93%, 53.52% and 27.71%, respectively. Moreover, taking the rule activation ratio into consideration, we can infer that activating too many rules that contain noisy and less-related rules to the input query data would degrade the performance of the EBRB classification system. The proposed rule activation method is able to select a tighter highly input-related subset for ER.

As seen in Fig. 6, concerning the increased accuracy, M-EBRB achieves the best performance in all three categories of data sets among the three rule activation methods. Additionally, we can determine that the three rule activation methods enable great progress in multi-class classification. The outcome implies that the multi-class classification is more sensitive to rule activation.

Wilcoxon's test on the results of the accuracy comparison concerning C-EBRB, D-EBRB, and N-EBRB are shown in Table 4.

From the corresponding p-value, we can infer that the differences are significant. The accuracy of the 4 EBRB methods shows that M-EBRB is better than the other methods.

The experimental results with respect to response time for the 12 data sets are shown in Table 5. The best performances are underlined.

From the perspective of response time, using C-EBRB, most of the extended belief rules need to be activated in the ER procedure, resulting in a severe computational cost and making the C-EBRB unusable in time-critical classification applications. The average response time of D-EBRB is 52.2 ms, making this method the most time consuming among the four methods. In particular, the response time is 393.5 ms when classifying the Yeast data set. This is because to activate a sufficient number of extended belief rules, a parameter-tuning procedure must be executed using D-EBRB, which increases both the classification accuracy and response time. Compared to C-EBRB and D-EBRB, N-EBRB greatly decreases the response time. In N-EBRB, the space is pre-divided into several parts by introducing lines that pass through the sample data, thus making the rule activation procedure quite efficient. However, as we can see in the response time of the Cancer data set, the construction of a divided space is

Table 5
Response time (ms) of EBRB methods.

Data set	No. of attributes	C-EBRB	D-EBRB	N-EBRB	M-EBRB
Iris	4	3.3	3.3	1.2	0.5
Seeds	7	9.5	9.5	3.4	0.7
Glass	9	9.3	18.7	4.2	1.8
Ecoli	7	11.9	17.8	4.9	1.3
Diabetes	8	12.3	14.3	7.1	1.5
Cancer	30	9.3	10.0	35.2	1.3
Mammographic	5	30.2	28.0	7.5	1.0
Banknote	4	71.0	73.1	1.1	1.0
Yeast	8	251.3	393.5	1.9	1.6
Titanic	3	1.4	8.7	2.0	0.9
Banana	2	2.7	13.9	2.4	1.9
Phoneme	5	4.6	34.5	8.4	2.6
Average		34.7	52.2	6.6	1.3

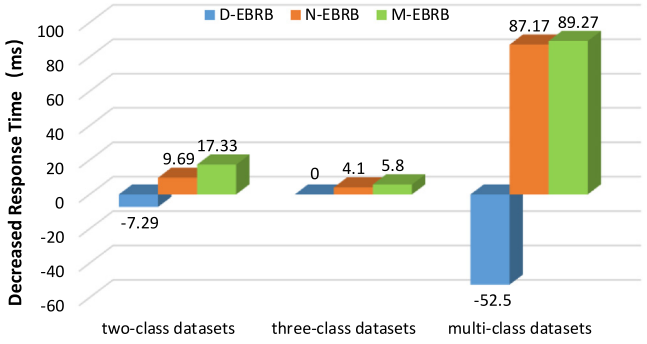

Fig. 7. Decreased response time of different rule activation methods compared to C-EBRB.

time intensive when addressing high-dimensional problems. The process of dividing the space is essentially a linear programming model. When classifying the Cancer data set, N-EBRB is the most time-consuming method among the four methods.

It can be observed in Fig. 7 that the response time is greatly reduced when using M-EBRB. The average response time is 1.3 ms for the 12 data sets. On the Banknote and Yeast data sets, M-EBRB has comparable performance to N-EBRB, which is much better than that of C-EBRB and D-EBRB. In addition, note that even though the response time is almost doubled when addressing large-scale data sets, e.g., the Banana and Phoneme data sets, the time consumption is also far lower than that of the other methods. The time complexity of M-EBRB is O(N), where N represents the scale of input. Specifically, the time complexity of M-EBRB is linearly increased with increasing input. Additionally, as noted in [31] (C-EBRB) [23] (D-EBRB) and [29] (N-EBRB), the time complexity of these 3 rule-based methods is O(N). However, the same time complexity does not mean the same computing time. As shown in Table 5 of our revised paper, the response time of our method is much shorter than the other 3 rule-based methods. This is because there is no programming or parameter-tuning scheme in M-EBRB. Additionally, in many cases, the sparsely distributed data may lead to Case 1 and Case 2, which would make the inference process quite fast. As shown in the bottom of Table 5, the average response times of C-EBRB, D-EBRB and N-EBRB are 25.7-, 39.2- and 5.1-times greater than that of M-EBRB, respectively. Thus, M-EBRB is more appropriate for addressing large-scale data classification problems.

As shown in the graph above, N-EBRB and M-EBRB decrease the response time for all three categories of data sets, especially for the multi-class classification. M-EBRB outperforms the other two rule activation methods. Additionally, the response time would increase when using DRA. For multi-class classification, the situation is severe and may be caused by the time-consuming parameter-tuning process.

Table 6
Wilcoxon's test of accuracy for comparison of M-EBEB with FRBCS under attribute noise ($\alpha = 0.05$).

Attribute noise level	z-value	*p*-value	Hypothesis
0%	−2.981	2.873 E-3	Rejected
5%	−2.824	4.742 E-3	Rejected
10%	−2.904	3.689 E-3	Rejected
15%	−2.903	3.702 E-3	Rejected
20%	−2.903	3.702 E-3	Rejected

Table 7
Wilcoxon's test of accuracy for comparison of M-EBEB with FRBCS under label noise ($\alpha = 0.05$).

Label noise level	z-value	*p*-value	Hypothesis
0%	−2.981	2.873 E-3	Rejected
5%	−2.981	2.873 E-3	Rejected
10%	−3.059	2.218 E-3	Rejected
15%	−3.059	2.218 E-3	Rejected
20%	−3.059	2.218 E-3	Rejected

4.5. Robustness evaluation

To verify the effectiveness of M-EBRB in handling noisy data, we conducted some experiments. Inspired by the techniques in [39], we generated noisy data for all the data sets in our paper. The noisy data can be divided into two parts: attribute noise and label noise. The level of attribute noise means the percentage of samples whose attribute values are randomly generated between the minimum values and maximum values of the corresponding attributes. Similarly, the noisy label data are randomly generated among the labels included in the data set. In comparison with M-EBRB, we used the fuzzy rule-based classification system using Chi's technique (FRBCS.CHI) [40]. The source code, which is in R, was downloaded from https://cran.r-project.org/web/packages/frbs/index.html, and some further information is shared at https://sci2s.ugr.es/fss. All the experimental outcomes are averaged from 20 repeated experiments (with different random seeds to select samples to generate rules). The attribute noise of *x*% means that we randomly select *x*% of all sample data and randomly generate the attribute values within the range of maximum and minimum values of the corresponding attribute. Similarly, the label noise is randomly generated from all the labels of the data set. The outcomes of the experiments under different noise levels are as follows:

In Fig. 8, the red solid lines are the trend of EBRB with MCDRA under different levels of attribute noise, and the red dotted lines are that under label noise. The blue solid lines and blue dotted lines are those of FRBCS under attribute noise and label noise, respectively. We can observe that except for on the Yeast data set, EBRB with MCDRA outperforms FRBCS.CHI under different attribute and label noise levels.

To illustrate the process of address noisy data using MCDRA, the data used in subsection 4.2 are adopted to generate the corresponding noisy data with 20% attribute noise and 20% label noise. With the same input as that in subsection 4.2, the rule activation results using data with 20% attribute noise and 20% label noise are as shown in Fig. 9 and Fig. 10, respectively.

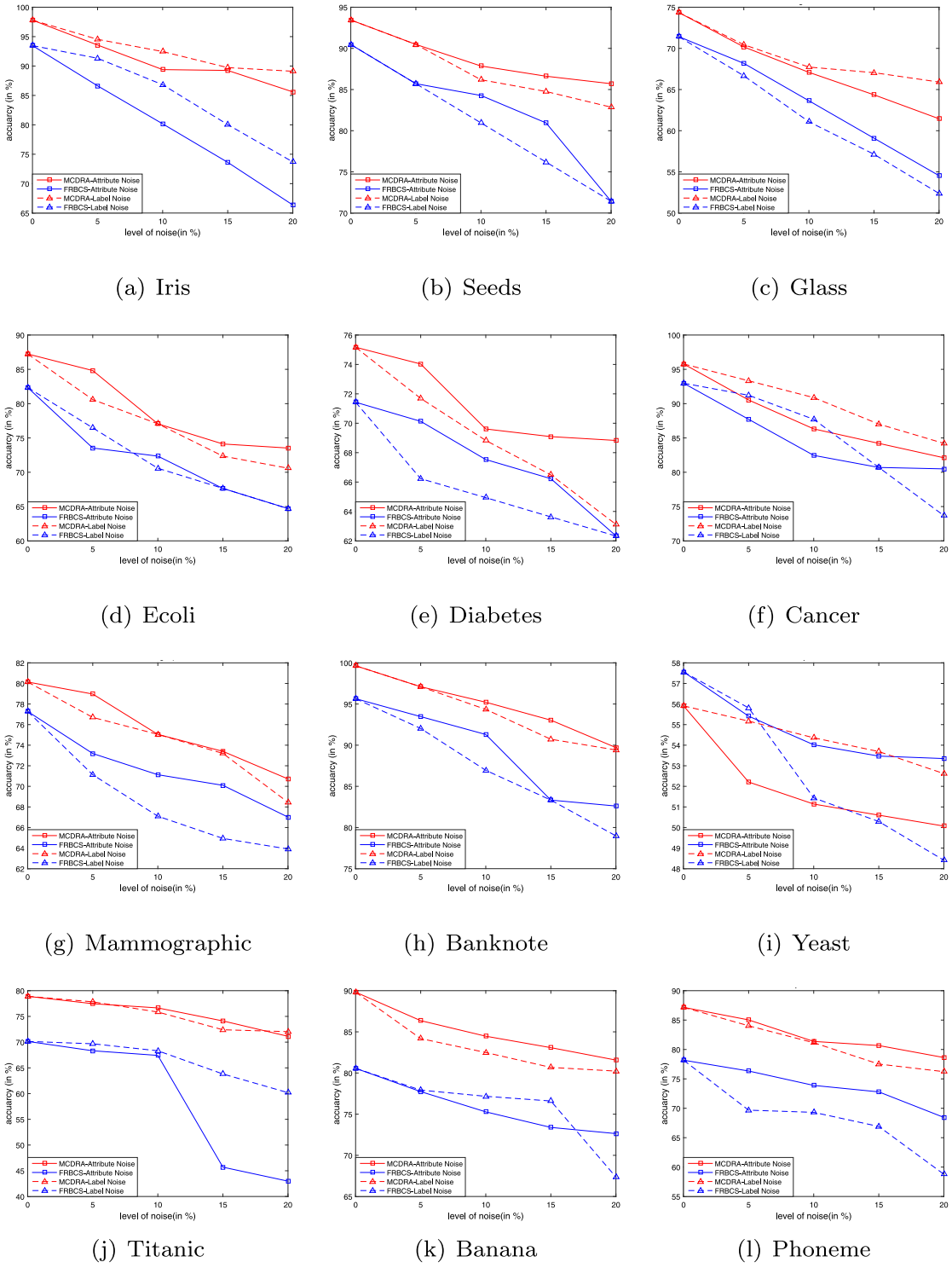

Fig. 8. Performance of EBRB with MCDRA and FRBCS under noisy data.

As shown in Figs. 9 and 10, the filtering procedure is evidentially influenced by the noisy data. However, the rule activation results are not strongly influenced compared with the result shown in Fig. 5, which is the result of the original data. The reason for the good performance of MCDRA is that the final activated rules are determined by the minimum distance between the input query data and the weighted centres, which is insensitive to noisy data.

The Wilcoxon's test of the results of the robustness evaluation comparison concerning M-EBRB and FRBCS are given in Tables 6 and 7:

The low p-value implies that the difference of M-EBRB and FRBCS concerning accuracy performance under various attribute noise levels and label noise levels is significant. Considering the better performance of M-EBRB, we can say that M-EBRB achieves better classification accuracy.

Table 8
RLA comparison under attribute noise (in %).

Attribute noise level	5%		10%		15%		20%	
	MCDRA	FRBCS	MCDRA	FRBCS	MCDRA	FRBCS	MCDRA	FRBCS
Iris	4.325153	7.317856	8.599182	14.229164	8.742331	21.236760	12.474438	28.993260
Seeds	3.178170	5.261413	5.981808	6.842047	7.308721	10.522825	8.282504	21.045650
Glass	5.621302	4.549909	9.776762	10.905782	13.448090	17.275654	17.361485	23.645527
Ecoli	2.773957	10.710383	11.668959	12.143291	15.038973	17.862781	15.715268	21.420765
Diabetes	1.516562	1.819964	7.396568	5.459891	8.088333	7.279854	8.434216	12.725745
Cancer	5.491179	5.657131	9.886209	11.314261	12.088945	13.207141	14.281240	13.432996
Mammographic	1.459945	5.292443	6.351385	7.958075	8.410282	9.290890	11.754430	13.289337
Banknote	2.558956	2.279143	4.445559	4.547831	6.633216	12.880293	9.974912	13.633037
Yeast	6.617779	3.735233	8.531569	6.150104	9.497407	7.105629	10.445359	7.314107
Titanic	1.824864	2.594810	2.864022	3.877958	6.070207	34.844597	9.859333	38.722555
Banana	3.852149	3.512474	5.956357	6.565719	7.503897	8.911506	9.185037	9.842373
Phoneme	2.454973	2.353543	6.653665	5.487337	7.456694	6.894346	9.796948	12.445638

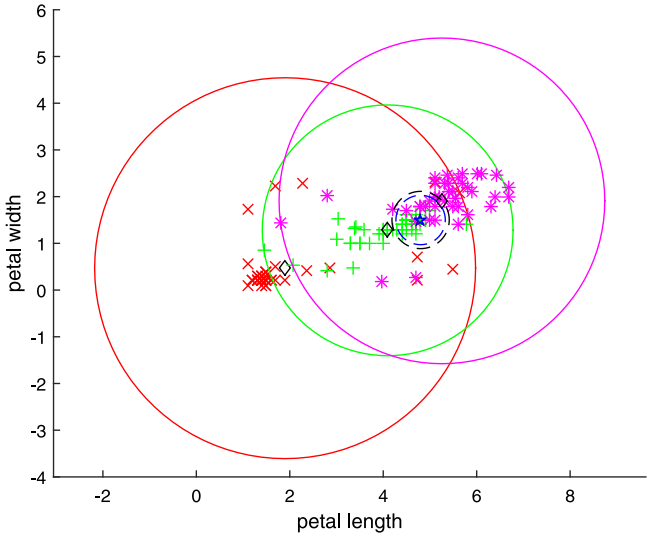

Fig. 9. Illustration of MCDRA under 20% iris attribute noise.

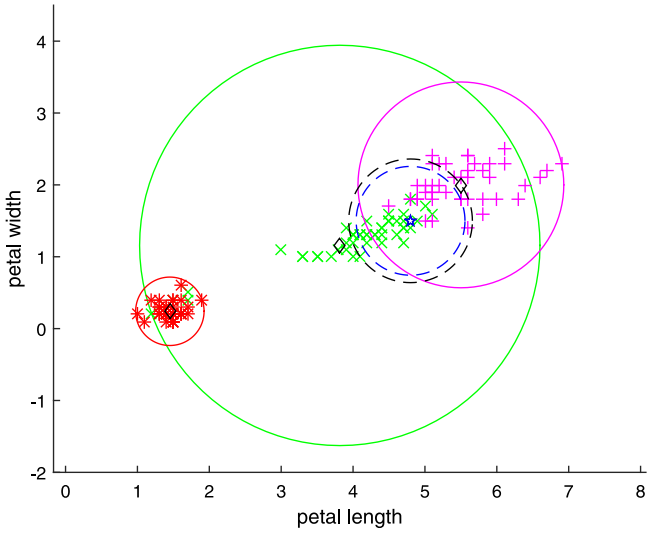

Fig. 10. Illustration of MCDRA under 20% iris label noise.

To verify the robustness of the two methods, the relative loss of accuracy (RLA) is utilized to measure the performance of FRBCS and EBRB with MCDRA.

$$RLA_{x\%} = \frac{Acc_{0\%} - Acc_{x\%}}{Acc_{0\%}} \tag{18}$$

where $RLA_{x\%}$ represents the relative loss of accuracy at the noise level of $x\%$. $Acc_{0\%}$ and $Acc_{x\%}$ are the accuracies of the classifier at noise levels of 0% and $x\%$, respectively. The experimental outcome can be seen in Tables 8 and 9.

We can infer that overall, MCDRA outperforms FRBCS concerning RLA. Note that FRBCS performs better in addressing attribute noise. For the condition with 20% noise, except for the Cancer and Yeast data sets with attribute noise, MCDRA demonstrated its capability in addressing noisy data.

4.6. Performance comparison of M-EBRB and machine learning algorithms

In this experiment, the performance and settings of M-EBRB are the same as those of the comparison experiment with the other rule activation methods. The machine learning algorithms are executed with the Waikato Environment for Knowledge Analysis (WEKA) software [41], which is developed by the Machine Learning Group from Waikato University. k-nearest neighbour (KNN), naive Bayes (NB), C4.5, support vector machine (SVM) and an artificial neural network are utilized to classify the 12 data sets used in our paper. For these algorithms, the settings are as follows: 5% of the samples in the training are set as k for KNN, the minimum number of instances per leaf is set as 2 for C4.5, the polynomial kernel is adopted in SVM and 50% of the sum of the number of attributes and classes is the number of hidden layers for the ANN. The accuracies of the aforementioned machine learning algorithms and M-EBRB are shown in Table 10. The best performances are underlined.

It can be observed that compared with the machine learning algorithms, M-EBRB produces satisfactory accuracy, achieving the best performance for Iris, Glass, Ecoli and Phoneme. Meanwhile, the proposed method achieves the second highest accuracy on Seeds and Yeast. It can be stated that M-EBRB outperforms the other algorithms on multi-class classification problems.

Wilcoxon's test on the result of the accuracy comparison among M-EBRB and the machine learning algorithms is shown in Table 11.

Based on Tables 10 and 11, ANN achieves the highest accuracy on the 4 data sets, and the rankings for SVM, NB and C4.5 are 2, 1 and 1, respectively. The hypothesis for the ANN is accepted, i.e., M-EBRB is comparable to the ANN. By comparing the accuracy, we can draw the same conclusion. The p-value and the accuracy of the other machine learning algorithms

Table 9
RLA comparison under label noise (in %).

Label noise level	5%		10%		15%		20%	
	MCDRA	FRBCS	MCDRA	FRBCS	MCDRA	FRBCS	MCDRA	FRBCS
Iris	3.333333	2.300203	5.460123	7.114582	8.251534	14.368246	8.885481	21.161870
Seeds	3.178170	5.261413	7.768860	10.522825	9.299090	15.784238	11.332263	21.045650
Glass	5.258203	6.663867	8.916084	14.475710	9.830554	20.005600	11.363636	26.669467
Ecoli	7.622650	7.140255	11.668959	14.292653	17.067859	17.850638	19.085282	21.432908
Diabetes	4.629506	7.279854	8.434216	9.085818	11.547160	10.919782	16.030331	12.739745
Cancer	2.568118	1.892880	5.125796	5.657131	9.155444	13.207141	12.088945	20.757152
Mammographic	4.292488	7.958075	6.351385	13.172878	8.659845	15.967909	14.586973	17.300725
Banknote	2.558956	3.784631	5.318615	9.095661	8.961365	12.880293	10.265931	17.428123
Yeast	1.323556	3.040306	2.772313	10.632384	3.970667	12.630299	5.884457	15.896456
Titanic	1.368648	0.655831	3.877836	2.594810	8.249905	9.024808	8.706121	14.200171
Banana	6.279225	3.289065	8.160766	4.219933	10.175907	4.927392	10.676909	16.395681
Phoneme	3.602157	10.872346	6.906046	11.345613	11.070322	14.415452	12.550189	24.814531

Table 10
Accuracy (%) comparison of M-EBRB and machine learning algorithms.

Data set	KNN	NB	C4.5	SVM	ANN	M-EBRB
Iris	96.67	96.00	96.00	96.67	97.33	97.80
Seeds	92.38	91.43	91.90	90.48	95.24	93.45
Glass	66.36	48.60	66.82	68.69	67.76	74.36
Ecoli	85.71	85.42	84.23	75.60	86.01	87.24
Diabetes	74.09	76.30	73.82	65.10	75.39	75.17
Cancer	95.96	92.97	93.32	62.74	96.31	95.79
Mammographic	78.46	78.36	82.10	79.29	80.96	80.14
Banknote	98.98	84.26	98.54	98.03	99.93	99.65
Yeast	58.22	57.61	55.39	43.26	59.03	58.62
Titanic	78.15	74.69	79.06	79.22	77.87	78.91
Banana	89.94	61.11	89.04	90.25	72.09	89.82
Phoneme	79.44	76.05	86.42	84.49	80.98	87.17

Table 11
Wilcoxon's test of accuracy for comparison of M-EBEB with machine learning algorithms ($\alpha = 0.05$).

Methods	z-value	*p*-value	Hypothesis
KNN	−2.824	4.742 E-3	Rejected
NB	−2.903	3.702 E-3	Rejected
C4.5	−2.353	1.860 E-2	Rejected
SVM	−2.824	4.700 E-3	Rejected
ANN	−1.020	0.3078	Accepted

indicate that M-EBRB outperforms the other algorithms. Essentially, there is no perfect algorithm for all data sets because the characteristics of the algorithms and data sets vary greatly. In summary, compared with machine learning algorithms, M-EBRB has comparative performance on classification problems.

4.7. Discussion on the performance of MCDRA

Using MCDRA, the rule activation is strengthened in the following aspects. First, we added a noise filtering procedure so that we could decrease the influence of input-unrelated sample data and substantially reduce the computing cost of the subsequent procedures in MCDRA. Second, we added a density adjusting factor, which helps query data in sparsely distributed areas activate more extended belief rules and those in densely distributed areas activate fewer results for the ER process. Finally, we manage to eliminate subjective information through the minimum centre distance. Through the above-mentioned procedures, using MCDRA, the rule activation for every input query data sample is unique and customized by the position and distribution of the whole data set; thus, the rule base is strengthened.

It is noticeable that, even with different working scheme, both the proposed MCDRA and DRA in [28] are distance-based rule activation methods. The case study shows that MCDRA achieves relatively better results concerning response time and classification accuracy, which are key indicators of rule activation methods. The inherent properties of MCDRA that contribute to the results may be that the filtering procedure greatly downsize computing cost of the selection process, which result in the short response time of MCDRA. However, for DRA no data preprocessing procedure is involved before the selection of activated belief rules. Moreover, the time-consuming parameter tuning process result in longer response time of DRA, while this factor in selection algorithm of MCDRA is obtained quite efficiently. For the classification accuracy, the distance adjustment factor of selection algorithm in MCDRA is determined by the distribution data set, which result in a proper number of activated rules. However, the range of this factor is given by experts, which may be not proper for some data sets.

Nevertheless, there are some shortcomings to our work. The rule activation ratio for multiple classes is much higher than that of two-class and three-class data sets. This is caused by the filtering mechanism of MCDRA. Additionally, the design of MCDRA can be used to automatically conduct experiments on data sets and ignore subjective information, which is good for large-scale and middle-scale data sets. When addressing very-small-scale data sets, e.g., 10 rules, expert knowledge may improve the performance of the classifier better.

5. Conclusion and future research

In our study, a new EBRB rule activation method is proposed to address classification problems. To verify the validity of M-EBRB, 12 data sets are tested using conventional EBRB, two EBRB rule activation methods, M-EBRB and machine learning algorithms. Moreover, the performance of M-EBRB and FRBCS under different noise levels is compared.

The contributions of our paper can be summarized as follows: First, M-EBRB greatly improves the performance of rule activation compared to C-EBRB, which faces the problem of activating nearly the entire extended belief rule base. Second, M-EBRB, in terms of accuracy and response time, outperforms the other two rule activation methods on most of the datasets used in this paper. In addition, M-EBRB managed to eliminate subjective information and time-consuming iteration and programming procedures. The experiments under different noise levels showed that M-EBRB is able to maintain a more stable performance compared to FRBCS.

Finally, in comparison with other state-of-the-art machine learning algorithms, M-EBRB achieves comparable performance to the ANN and outperforms the other algorithms.

MCDRA works well in downsizing the extended belief rule base while obtaining better performance. In future research, parameter optimization will be one challenge in improving the performance of M-EBRB when applied to classification problems, especially when addressing large-scale and high-dimensional problems. We will investigate in the influence of attribute weight and rule weight on the performance of M-EBRB.

Declaration of competing interest

No author associated with this paper has disclosed any potential or pertinent conflicts which may be perceived to have impending conflict with this work. For full disclosure statements refer to https://doi.org/10.1016/j.asoc.2020.106214.

CRediT authorship contribution statement

Haizhen Zhu: Conceptualization, Methodology, Software, Writing - original draft, Writing - review & editing. **Mingqing Xiao:** Supervision, Validation. **Longhao Yang:** Investigation, Methodology. **Xilang Tang:** Formal analysis, Data curation. **Yajun Liang:** Software, Visualization. **Jianfeng Li:** Software, Validation.

References

[1] L. Lin, K.Z. Wang, D.Y. Meng, W.M. Zuo, L. Zhang, Active self-paced learning for cost-effective and progressive face identification, IEEE Trans. Pattern Anal. Mach. Intell. PP (99) (2018) 7–19.

[2] J.A. Sáez, B. Krawczyk, M. Woźniak, On the influence of class noise in medical data classification: treatment using noise filtering methods, Appl. Artif. Intell. 30 (6) (2016) 20.

[3] Y.W. Ge, M.Q. Xiao, Z. Yang, L. Zhang, Y.J. Liang, A hybrid hierarchical fault diagnosis method under the condition of incomplete decision information system, Appl. Soft Comput. 73 (2018) 350–365.

[4] A.S. Ghareb, A.A. Bakar, A.R. Hamdan, Hybrid feature selection based on enhanced genetic algorithm for text categorization, Expert Syst. Appl. 49 (2016) S0957417415007952.

[5] A. Zehtabian, A. Nazari, H. Ghassemian, M. Gribaudo, Adaptive Restoration of Multispectral Datasets used for SVM classification, Eur. J. Remote Sens. 48 (48) (2015) 183–200.

[6] J. Maillo, S. Ramírez, I. Triguero, F. Herrera, KNN-IS: An Iterative Spark-based design of the k-Nearest Neighbors Classifier for Big Data, Knowl.-Based Syst. (2016) S0950705116301757.

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

[8] Z.J. Zhou, C.H. Hu, J.B. Yang, D.L. Xu, D.H. Zhou, Online updating belief rule based system for pipeline leak detection under expert intervention, Expert Syst. Appl. 36 (4) (2009) 7700–7709.

[9] L.M. Jiao, T. Denoeux, P. Quan, A hybrid belief rule-based classification system based on uncertain training data and expert knowledge, IEEE Trans. Syst. Man Cybern. Syst. PP (99) (2016) 1–13.

[10] Z.G. Liu, Q. Pan, J. Dezert, G. Mercier, Y. Liu, Fuzzy-belief K-nearest neighbor classifier for uncertain data, in: International Conference on Information Fusion, 2014.

[11] J.A. Sanz, A. Fernández, H. Bustince, F. Herrera, IVTURS: A linguistic fuzzy rule-based classification system based on a new interval-valued fuzzy reasoning method with tuning and rule selection, IEEE Trans. Fuzzy Syst. 21 (3) (2013) 399–411.

[12] J.B. Yang, J. Liu, D.L. Xu, J. Wang, H.W. Wang, Optimization models for training belief-rule-based systems, IEEE Trans. Syst. Man Cybern. A 37 (4) (2007) 569–585.

[13] Z.J. Zhou, C.H. Hu, D.L. Xu, J.B. Yang, D.H. Zhou, Bayesian reasoning approach based recursive algorithm for online updating belief rule based expert system of pipeline leak detection, Expert Syst. Appl. 38 (4) (2011) 3937–3943.

[14] X.L. Tang, M.Q. Xiao, Y.J. Liang, H.Z. Zhu, J.F. Li, Online updating belief-rule-base using Bayesian estimation, Knowl.-Based Syst. 171 (2019) 93–105, http://dx.doi.org/10.1016/j.knosys.2019.02.007.

[15] X.B. Xu, J. Zheng, J.B. Yang, D.L. Xu, Y.W. Chen, Data classification using evidence reasoning rule, Knowl.-Based Syst. 116 (C) (2017) 144–151.

[16] L.M. Jiao, P. Quan, T. Denœux, L. Yan, X.X. Feng, Belief rule-based classification system: Extension of FRBCS in belief functions framework, Inform. Sci. 309 (2015) 26–49.

[17] M. Rey, M. Galende, M. Fuente, G. Sainz-Palmero, Multi-objective based Fuzzy Rule Based Systems (FRBSs) for trade-off improvement in accuracy and interpretability: A rule relevance point of view, Knowl.-Based Syst. 127 (2017) 67–84.

[18] L.L. Chang, Z.J. Zhou, Y. You, L.H. Yang, Z.G. Zhou, Belief rule based expert system for classification problems with new rule activation and weight calculation procedures, Inform. Sci. 336 (C) (2016) 75–91.

[19] J. Casillas, M.J.D. Jesus, F. Herrera, Genetic feature selection in a fuzzy rule-based classification system learning process for high-dimensional problems, Inform. Sci. 136 (1) (2001) 135–157.

[20] L.L. Chang, Z. Yu, J. Jiang, X.H. Li, Structure learning for belief rule base expert system: A comparative study, Knowl.-Based Syst. 39 (2) (2013) 159–172.

[21] L.H. Yang, J. Liu, Y.M. Wang, L. Martinez, A micro-extended belief rule-based system for big data multiclass classification problems, IEEE Trans. Syst. Man Cybern. Syst. PP (99) (2018) 1–21, http://dx.doi.org/10.1109/TSMC.2018.2872843.

[22] M. Espinilla, J. Medina, A. Calzada, J. Liu, L. Martínez, C. Nugent, Optimizing the configuration of an heterogeneous architecture of sensors for activity recognition, using the extended belief rule-based inference methodology, Microprocess. Microsyst. (2016) S0141933116302824.

[23] A. Calzada, J. Liu, W. Hui, A. Kashyap, Dynamic rule activation for Extended Belief Rule Bases, in: International Conference on Machine Learning and Cybernetics, 2014.

[24] L.H. Yang, Y.M. Wang, Y.X. Lan, L. Chen, Y.G. Fu, A data envelopment analysis (DEA)-based method for rule reduction in extended belief-rule-based systems, Knowl.-Based Syst. 123 (C) (2017) 174–187.

[25] R.Y. Yu, L.H. Yang, Y.G. Fu, Data driven construction and inference methodology of belief rule-base, J. Comput. Appl. 34 (8) (2014) 2155–2160, http://dx.doi.org/10.11772/j.issn.1001-9081.2014.08.2155.

[26] L.H. Yang, Y.M. Wang, Q. Su, Y.G. Fu, K.S. Chin, Multi-attribute search framework for optimizing extended belief rule-based systems, Inform. Sci. 370–371 (2016) 159–183, http://dx.doi.org/10.1016/j.ins.2016.07.067.

[27] L.H. Yang, Y.M. Wang, Y.G. Fu, A consistency analysis-based rule activation method for extended belief-rule-based systems, Inform. Sci. 445–446 (5) (2018) 50–65, http://dx.doi.org/10.1016/j.ins.2018.02.059.

[28] A. Calzada, J. Liu, H. Wang, A. Kashyap, A new dynamic rule activation method for extended belief rule-based systems, Knowl. Data Eng. IEEE Trans. 27 (4) (2015) 880–894, http://dx.doi.org/10.1109/tkde.2014.2356460.

[29] L.H. Yang, J. Liu, Y.M. Wang, L. Martínez, Extended belief-rule-based system with new activation rule determination and weight calculation for classification problems, Appl. Soft Comput. (2018) http://dx.doi.org/10.1016/j.asoc.2018.08.004.

[30] J.B. Yang, J. Liu, W. Jin, H.S. Sii, H.W. Wang, Belief rule-base inference methodology using the evidential reasoning Approach-RIMER, IEEE Trans. Syst. Man Cybern. A 36 (2) (2006) 266–285, http://dx.doi.org/10.1109/TSMCA.2005.851270.

[31] J. Liu, L. Martinez, A. Calzada, W. Hui, A novel belief rule base representation, generation and its inference methodology, Knowl.-Based Syst. 53 (53) (2013) 129–141, http://dx.doi.org/10.1016/j.knosys.2013.08.019.

[32] L. Jiao, X. Geng, Q. Pan, Compact belief rule base learning for classification with evidential clustering, Entropy 21 (5) (2019) 443.

[33] Z.G. Liu, P. Quan, J. Dezert, G. Mercier, Credal classification rule for uncertain data based on belief functions, Pattern Recognit. 47 (7) (2014) 2532–2541, http://dx.doi.org/10.1016/j.patcog.2014.01.011.

[34] H. Ma, H. Yi, H. Shi, Fault detection and identification based on the neighborhood standardized local outlier factor method, Ind. Eng. Chem. Res. 52 (6) (2013) 2389–2402.

[35] Z. Ataser, F.N. Alpaslan, Self-Adaptive Negative Selection Using Local Outlier Factor, 2013.

[36] L.H. Yang, J. Liu, Y.M. Wang, L. Martínez, New activation weight calculation and parameter optimization for extended belief rule-based system based on sensitivity analysis, Knowl. Inf. Syst. (4) (2018) 1–42, http://dx.doi.org/10.1007/s10115-018-1211-0.

[37] L. Jiao, X. Geng, Q. Pan, A compact belief rule-based classification system with evidential clustering, in: International Conference on Belief Functions, Springer, 2018, pp. 137–145.

[38] UCI, Repository of machine learning databases, 2019, https://archive.ics.uci.edu/ml/datasets.html. (Accessed 2 April 2019).

[39] J. Saez, J. Luengo, F. Herrera, On the suitability of fuzzy rule-based classification systems with noisy data, IEEE Trans. Fuzzy Syst. (2012).

[40] Z. Chi, H. Yan, T. Pham, Fuzzy Algorithms: with Applications to Image Processing and Pattern Recognition, Vol. 10, World Scientific, 1996.

[41] G. Holmes, A. Donkin, I.H. Witten, WEKA: A machine learning workbench, in: Conference on Intelligent Information Systems, 2002.