Information Fusion 124 (2025) 103386

Contents lists available at ScienceDirect

Information Fusion

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

Full length article

Association-based concept-cognitive learning for classification: Fusing knowledge with distance metric learning

Chengling Zhang [a], Guangming Xue [b], Weihua Xu [a,*], Huilai Zhi [c], Yinfeng Zhou [d], Eric C.C. Tsang [e]

[a] College of Artificial Intelligence, Southwest University, Chongqing, 400715, China
[b] College of Information Engineering, Nanning University, Nanning, 530200, China
[c] College of Mathematics and Computer Science, Quanzhou Normal University, Quanzhou, 362000, China
[d] School of Mathematics and Statistics, Shaanxi Normal University, Xian, 710119, China
[e] School of Computer Science and Engineering, Macau University of Science and Technology, Taipa, Macao Special Administrative Region of China

ARTICLE INFO

Keywords:
Associative learning
Concept-cognitive learning
Distance metric learning
Fuzzy formal context
Classification

ABSTRACT

Concept-cognitive learning, which emphasizes the representation and learning of knowledge incorporated within data, has yielded excellent results in classification research. However, learning concepts from a high-dimensional dataset is a time-consuming and complex process, which increases the extraction of redundant information and leads to poor classification task. Most existing neighborhood concept generated by neighborhood similarity granule use a single predefined distance function and ignore the decision labels, which lead to the fact that the learned distance function is not optimal. Moreover, current concept-cognitive learning methods do not fully utilize the advantages of granular concept and neighborhood concept, resulting in weak interpretability. To address these issues, we introduce a novel association-based concept-cognitive learning method with distance metric learning for knowledge fusion and concept classification. To be concrete, to decrease the dimensionality of dataset and remove the interfering information, the representative attribute set from attribute clusters based on correlation coefficient matrix is firstly discussed. Subsequently, neighborhood similarity granules based on distance metric learning are used to construct fuzzy concepts. To obtain fuzzy concept of maximum contribution, we present a valid fuzzy concept associative space related to clues in the human brain. Furthermore, a mechanism of fuzzy concept-cognitive associative learning with distance metric learning (FCADML) model is proposed, which aims to achieve concept clustering and class prediction by fusing objects and attributes within fuzzy concepts. Finally, we perform a classification performance evaluation on thirteen datasets which verify that the feasibility and efficiency of the proposed learning mechanism.

1. Introduction

Cognitive computing [1], as an important part of artificial intelligence, can simulate complex human thinking process to achieve cognitive intelligence, which is used to handle perception, concepts learning and extracting, memory and experienced knowledge acquisition and updating [2,3]. In fact, as the fundamental unit of human cognition, concept constructs a mapping between reality and mental word, and explores the nature of development of things. Concept typically composes of extent and intent, where extent is the collection of instances or objects represented by a concept, while intent is the set of attributes or properties that concept possesses [4,5]. So far, the integration of concept learning with philosophy, psychology, mathematics and other disciplines has been extensively utilized in diverse application fields including rule extraction [6,7], dynamic updating [8], object classification [9,10], etc.

With the rapid development and advancement in information technology, big data is growing explosively. It has the characteristics of high redundancy and high noise which increase the complexity of information recognition. To enhance the ability of machine learning models to identify data patterns, it is necessary to remove redundant and interfering information from data before training models. Currently, knowledge discovery and feature selection relying on rough set theory (RST) [11] and formal concept analysis (FCA) [4] have attracted the interests of many readers. To find attribute subsets with high separability and faster computational efficiency, it is essential to combine the dependence and structure of attribute subsets for attribute reduction. Prasad et al. [12] proposed a Bayesian rough set model

* Corresponding author.
E-mail addresses: chengling_zhang@126.com (C. Zhang), xueguangming0417@126.com (G. Xue), chxuwh@gmail.com (W. Xu), zhihuilai@126.com (H. Zhi), j_jenifer@126.com (Y. Zhou), cctsang@must.edu.mo (E.C.C. Tsang).

https://doi.org/10.1016/j.inffus.2025.103386
Received 19 March 2025; Received in revised form 11 May 2025; Accepted 2 June 2025
Available online 22 June 2025

which can solve uncertainty, classification and multi-decision problems. Yuan et al. [13] introduced an innovative uncertainty measure based on Zentropy and designed a feature selection approach by using the granule structure in the decision information system. Focusing on FCA, Shao et al. introduced the attribute reduction methods for maintaining granular concept and lattice structure from the perspective of reducing attribute in [14,15]. It should be noted that the dimensionality of data can be reduced through reduction algorithms for eliminating redundant information and improving data quality. However, traditional reduction methods may suffer from overfitting in the face of high-dimensional data, resulting in unreliable results. Hence, how to reduce the dimensionality of data to enhance classification performance in cognitive computing is the basic motivation of this paper.

With the development of the integration of cognitive computing and FCA, a novel research subject named concept-cognitive learning (CCL) has brought about new advancements. CCL is an interdisciplinary research field that studies cognition of things and acquisition of knowledge through concepts [16,17]. CCL recognizes concepts by specific cognitive models that simulate the behavior of the human brain from given clue and reveal the systematic patterns. Recently, some groundbreaking results have been achieved, such as Wille's formal concept [5], fuzzy concept [14,15,18], three-way dual concept [19], two-way concept leaning [20,21]. Besides, Yao et al. [17] introduced an attractive method of concept learning with the viewpoint of cognitive informatics and granular computing. To obtain sufficient and necessary concepts, Xu et al. [20] discussed a concept learning transformation method that combines granular computing with two-way learning. Considering that concepts change with the increase of information granules, Xu et al. [21] further investigated a dynamic two-way concept-cognitive learning from a fuzzy progressive learning perspective. Indeed, precise cognition of concept is usually difficult to achieve, so Li et al. [22] developed a CCL mechanism combined upper and lower approximations from philosophy and cognitive psychology.

In addition, up to now, models that integrate concept cognition with machine learning have attracted much attention. For instance, in a series of seminal articles by Mi et al. concept cognition was first combined with machine learning to present an incremental CCL model [23], fuzzy-based concept learning classification model [24] and concurrent CCL model [25]. On this basis, papers [10,26–28] successively put forward several CCl classification algorithms based on weighted granular concepts, fuzzy granular concepts and interval granular concepts. The above-mentioned models all have in common that the concept spaces are constructed through different forms of granular concept. Moreover, Yuan et al. [9] studied an incremental learning mechanism through the progressive fuzzy three-way concept. Next, to address high-dimensional data, Guo et al. [29] investigated a CCL system for tumor diagnosis to improve the interpretability and universality. Considering the accumulation and forgetting of knowledge in dynamic environment, Guo et al. [30] designed a memory-based CCL method by combining concept recalling and concept forgetting. From the above statement, we can find that neighborhood granules mainly rely on different similarity metrics (e.g., Euclidean distance or cosine theorem), and then concept spaces are obtained through fuzzy operators, which are used for classification problem. Nevertheless, there exist some limitations about the above techniques, which are manifested in: 1) There is no research indicating that granular concept or neighborhood concept induced by neighborhood similarity granule is more beneficial for representing concept learning and improving classification performance. 2) The existing CCL methods do not fully utilize the advantages of combining granular concept and neighborhood concept, resulting in weak interpretability. Toward this end, considering both granular concept and neighborhood concept to improve classification accuracy is another motivation for this paper.

Distance metric learning (DML) is brought forward by Xing et al. [31] to measure the similarity between samples. It has been extensively utilized in many funny machine learning tasks, such as classification [10,13,32], information retrieval [31,33], and bioinformatics [34]. DML is to compute the distance or similarity between diverse samples such that samples within the same decision class have a smaller distance than samples from the opposite decision class, and vice versa. The Euclidean distance metric is prevalent distance metric, which is generally used to measure the distance between two samples. However, this distance metric could be unsuitable for every application area since it only considers the linear distance between samples and ignores the correlation between attributes. It is worth stressing that the desirable distance metric should maintain the similarity relationship in dataset, that is, samples with high similarity should exhibit proximity, while dissimilar samples should maintain significant separation. For this problem, we aim to characterize the distance metric through considering conditional attributes and decision classes so that samples within the same class are more similar than those from different classes, thereby constructing an optimal neighborhood similarity granule. This is another motivation for this paper.

To counter the above-mentioned issues, we introduce an innovative association-based concept-cognitive learning method with distance metric learning for knowledge fusion and concept classification. Concretely, the representative attribute set via attribute clusters is discussed to reduce the dimensionality of data. It is known that associative learning can enable individuals to grasp concepts related to cues, which is highly consistent with the cognitive process in human brain cognition that triggers association and constructs concept through cues. Subsequently, to collect concepts with high contribution degree, we explore a fuzzy concept-cognitive associative learning mechanism for classification. The flowchart of the put-forward model FCADML is presented in Fig. 1. In summary, this article has the following contributions.

- In high-dimensional data mining, according to the correlation coefficient matrix, the representative attribute set based on attribute clusters is crucial to remove interfering information. This provides a new method for decreasing the computational cost and improving the operational efficiency.
- The decision information of samples needs to be considered to learn a suitable distance metric when granulating samples, which can not only improve the discrimination ability of similar relationship but also decrease the uncertainty of data, thereby enhance the accuracy and reliability of data analysis and processing.
- During the process of associative learning, the richness and flexibility of association are strengthened with the help of granular concept and neighborhood concept. Thus, fuzzy concept closest to clue is learned, so as to better meet the actual needs.

The subsequent arrangements of the article are organized as follows. Some fundamental notations of fuzzy formal concept analysis and DML are recalled in Section 2. Section 3 introduces a novel concept associative learning with DML. The results of our experimental evaluation appear in Section 4. At last, we summarize this work and provide recommendations for future research in Section 5.

2. Preliminaries

This section illustrates a brief overview of fuzzy formal concept analysis and distance metric learning. Further details are available in [18,31,35].

2.1. Fuzzy formal concept analysis

Given a universe U, and a fuzzy set $\tilde{F}$ of U is measured as a mapping function $\tilde{F}(\cdot) : U \to [0,1]$. Given each $u \in U$, the value $\tilde{F}(u)$ is interpreted as the fuzzy membership degree of u belonging to $\tilde{F}$. Let $\mathcal{F}(U)$ stand for the family of all fuzzy sets on U.

Assume that M and N are two fuzzy sets over U. If $\tilde{M}(u) \leq \tilde{N}(u)$, $u \in U$, then M is considered a subset of N, that is, $M \subseteq N$. Specifically, the set of all crisp sets on U is marked as $\mathcal{P}(U)$.

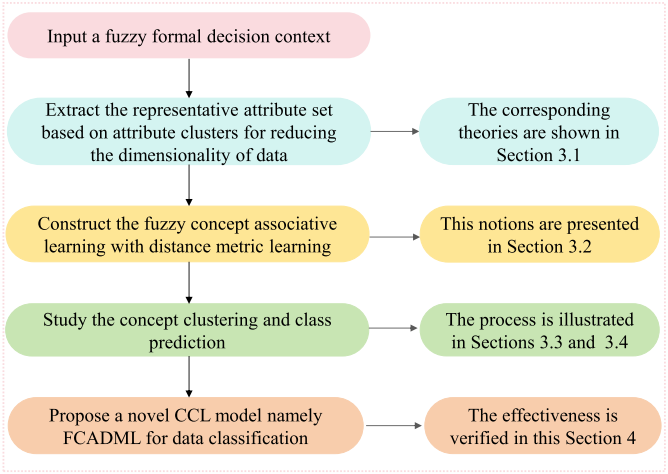

Fig. 1. Flowchart of model FCADML.

A fuzzy formal context is referred to as a triplet $(U, A, \tilde{R})$, where $U = \{u_1, u_2, \ldots, u_n\}$ is the set of objects and $A = \{a_1, a_2, \ldots, a_m\}$ is the set of attributes. Subsequently, $\tilde{R}$ represents a fuzzy relation between U and A (that is, $\tilde{R} : U \times A \to [0,1]$), and every $\tilde{R}(u,a)$ means the membership degree of object u to attribute a.

In a fuzzy formal context $(U, A, \tilde{R})$, given $Y \subseteq U$ and $\tilde{C} \in \mathcal{F}(A)$, two operators $\tilde{T} : \mathcal{P}(U) \to \mathcal{F}(A)$ and $H : \mathcal{F}(A) \to \mathcal{P}(U)$ are given by:

$$\begin{aligned} &\tilde{T}(Y)(a) = \bigwedge_{u \in Y} \tilde{R}(u,a), a \in A, \\ &H(\tilde{C}) = \left\{ u \in U : \forall a \in A, \tilde{C}(a) \le \tilde{R}(u,a) \right\}, \end{aligned} \tag{1}$$

where a pair $(Y, \tilde{C})$ is called fuzzy concept satisfying $\tilde{T}(Y) = \tilde{C}$ and $H(\tilde{C}) = Y$. Generally speaking, Y is extent and $\tilde{C}$ is intent of fuzzy concept.

In addition, with respect to two fuzzy formal contexts $(U, A, \tilde{R})$ and (U, D, J), where $\tilde{R} : U \times A \to [0,1]$ and $J : U \times D \to \{0,1\}$. Then the quintuple $(U, A, \tilde{R}, D, J)$ is called a fuzzy formal decision context (for short FFDC) where $A \cap D = \emptyset$ with A representing the conditional attribute set and D representing the decision attribute set. Then $U/D = \{U^{d_1}, U^{d_2}, \ldots, U^{d_t}\}$ is a decision division by D with $U = U^{d_1} \cup U^{d_2} \cup \cdots \cup U^{d_t}$ and $U^{d_1} \cap U^{d_2} \cap \cdots \cap U^{d_t} = \emptyset$.

2.2. Distance metric learning

Most existing sample similarities are characterized by distance metrics, such as Manhattan distance, Euclidean distance, Chebychev distance, etc. A distance function is denoted as

$$D_A(u,v) = \left(\sum_{i=1}^{m} |f(u,a_i) - f(v,a_i)|^p\right)^{1/p},$$

where u and v are two objects in an m-dimensional vector $A = \{a_1, a_2, \ldots, a_m\}$, and $f(u,a_i)$ and $f(v,a_i)$ are the values of objects u and v in the ith dimension a_i. To the best of our knowledge, these distance metrics fail to take into account the decision labels of samples. In fact, if the distance between two objects with different label is very small, it is easy to cause misclassification of objects. Therefore, given the decision information of the object, metric learning is to learn a valid distance function in which the nearest neighbors belonging to the same decision label are closely held and vice versa. Most researches focus on the Mahalanobis distance since it can be easily optimized by deriving a convex function that ensuring the global optimum [35,36]. Then a convex objective function for distance metric was proposed [31] and its form is given by as follows:

$$d_Q(u,v) = \sqrt{(Q_u - Q_v)^T M (Q_u - Q_v)}, \tag{2}$$

where $Q_u = (f(u,a_1), f(u,a_2), \ldots, f(u,a_q))^T$ is the vector of feature values in a feature subsect $Q = \{a_1, a_2, \ldots, a_q\}$. The learning form is parameterized by matrix M which is required to be positive semi-definite. Therefore, we can formulate a loss minimization problem as the function of the learn metric M.

The neighborhood of object is a collection of similar objects. In such case, the neighborhood of an object is the collection of similar objects that have the same decision label as the object attached and is defined as $d_Q(u,z) \le d_Q(u,v) + 1$, in which objects u and v are assigned to the same decision label, and u and z belong to different decision labels. Then the loss function consists of two terms induced by Eqs. (3) and (4), where Eq. (3) is used to pull those objects with the same decision label closer, and Eq. (4) is to push away those objects with different decision labels. That is, the form of problem is converted as follows:

$$pull(M) = \sum_{v \in sk_Q(u)} d_Q(u,v), \tag{3}$$

$$push(M) = \sum_{v \in sk_Q(u)} \sum_{z \in k_Q(u) - sk_Q(u)} [1 + d_Q(u,v) - d_Q(u,z)]_+, \tag{4}$$

where $[\cdot]_+ = max(\cdot, 0)$. $sk_Q(u)$ represents object u has k nearest neighbors that belong to the same decision label. $k_Q(u)$ means that u has k nearest neighbors in dataset. Subsequently, these two terms are integrated into a loss function:

$$\begin{aligned} &\min_M \{(1-w)pull(M) + wpush(M)\}, \\ &s.t\ M \ge 0. \end{aligned} \tag{5}$$

where w is a constant. The minimization solution of loss function can be derived using gradient descent. It should be noted that this distance metric takes into account the decision label so that samples from the same decision label are close to each other, while samples from different decision labels are far from each other. Then assume M is a diagonal matrix, diagonal elements can serve as feature weights so as to evaluate feature importance.

3. Concept associative learning with distance metric learning

For ease of understanding below, we first provide explanations of some notations, which is illustrated in Table 1.

3.1. Attribute clusters for representative attributes

In the big data era, data has an explosive growth, but in fact, features with high correlation play a crucial part in the information process. Thus, it is essential to reduce the redundant or less relevant features. In this subsection, we establish a method to extract representative attributes from attribute clusters.

Definition 1. Given a fuzzy formal context $(U, A, \tilde{R})$, for any $a_k, a_j \in A$, the correlation coefficient between a_k and a_j is defined as follows:

$$r(a_k, a_j) = \frac{\sum_{i=1}^{n} |\tilde{R}(u_i,a_k) - \bar{\tilde{R}}(a_k)| \, |\tilde{R}(u_i,a_j) - \bar{\tilde{R}}(a_j)|}{\sqrt{\sum_{i=1}^{n} (\tilde{R}(u_i,a_k) - \bar{\tilde{R}}(a_k))^2} \sqrt{\sum_{i=1}^{n} (\tilde{R}(u_i,a_j) - \bar{\tilde{R}}(a_j))^2}}, \tag{6}$$

where $\tilde{R}(u_i,a_k)$ and $\bar{\tilde{R}}(a_k)$ represent the membership degree of object u_i to attribute a_k and the mean value of attribute a_k with respect to the object set U, respectively. That is, $\bar{\tilde{R}}(a_k) = \frac{1}{n}\sum_{i=1}^{n} \tilde{R}(u_i,a_k)$. It is widely known that $r(a_k,a_j) \le 1$. Especially, if $\tilde{R}(u_i,a_k) = \tilde{R}(u_i,a_j)$ for any $u_i \in U$, then $r(a_k,a_j) = 1$ holds. Otherwise, $r(a_k,a_j) < 1$. The collection of correlation coefficient for any pair of attributes will construct a correlation coefficient matrix and is denoted as S, where $S(k,j) = r(a_k,a_j)$ for $a_k, a_j \in A$. It is obvious that the above matrix satisfies reflexivity and symmetry.

Example 1. A FFDC is described in Table 2 containing $U = \{u_1, u_2, \ldots, u_{13}\}$ and $A = \{a_1, a_2, \ldots, a_7\}$. The decision attribute d divides the

Table 1
The interpretation of notations.

Notation	Interpretation	Notation	Interpretation
U	A set of objects	U^{d_k}	A decision class
PR_i^δ	Attribute similarity path of ith path	PR^δ	All attribute similarity paths
CL_i	ith attribute cluster	$C\mathcal{L}$	All attribute clusters
$Q(CL_i)$	Representative attribute in CL_i	Q	Representative attribute set
$N(u)$	Neighborhood similarity granule of object u	$\mathcal{N}^{d_k}$	All neighborhood similarity granules in U^{d_k}
$MC^{d_k}(u)$	Maximum clue of object u in U^{d_k}	$\mathcal{MC}^{d_k}$	The set of all maximum clues in U^{d_k}
$FAS^{d_k}(MC^{d_k}(u))$	Fuzzy concept associative subspace of $MC^{d_k}(u)$ in U^{d_k}	$EAS^{d_k}(MC^{d_k}(u))$	Efficient fuzzy concept associative subspace of $MC^{d_k}(u)$ in U^{d_k}
$VCAS^{d_k}$	Valid fuzzy concept associative subspace in U^{d_k}	$VCAS^d$	Valid fuzzy concept associative space

Table 2
A fuzzy formal decision context.

U	a_1	a_2	a_3	a_4	a_5	a_6	a_7	d
u_1	0.73	0.33	0.59	0.53	0.50	0.40	0.65	1
u_2	0.10	0.79	0.33	0.50	0.45	0.04	0.10	1
u_3	0.44	0.50	0.09	0.97	0.33	0.46	0.67	1
u_4	0.37	0.43	0.33	0.15	0.53	0.30	0.73	1
u_5	0.32	0.42	0.28	0.85	0.14	0.31	0.53	1
u_6	0.69	0.85	0.84	0.92	0.03	0.57	0.88	1
u_7	0.96	0.22	0.88	0.45	0.17	0.45	0.39	1
u_8	0.20	0.27	0.31	0.34	0.55	0.21	0.73	2
u_9	0.61	0.01	0.40	0.52	0.90	0.20	0.02	2
u_{10}	0.45	0.44	0.87	0.26	0.58	0.75	0.70	2
u_{11}	0.67	0.06	0.67	0.64	0.13	0.88	0.79	2
u_{12}	0.07	0.37	0.65	0.36	0.05	0.17	0.27	2
u_{13}	0.25	0.61	0.43	0.46	0.13	0.09	0.49	2

whole set of objects into two categories where $U_1^d = \{u_1, u_2, u_3, u_4, u_5, u_6, u_7\}$ and $U_2^d = \{u_8, u_9, u_{10}, u_{11}, u_{12}, u_{13}\}$. The correlation coefficient matrix S is shown as:

$$S = \begin{bmatrix} 1.0000 & 0.7424 & 0.6987 & 0.4730 & 0.7453 & 0.6703 & 0.7681 \\ 0.7424 & 1.0000 & 0.6565 & 0.4736 & 0.7943 & 0.7899 & 0.8988 \\ 0.6987 & 0.6565 & 1.0000 & 0.8837 & 0.7064 & 0.7055 & 0.7124 \\ 0.4730 & 0.4736 & 0.8837 & 1.0000 & 0.6154 & 0.5687 & 0.5737 \\ 0.7453 & 0.7943 & 0.7064 & 0.6154 & 1.0000 & 0.7486 & 0.8790 \\ 0.6703 & 0.7899 & 0.7055 & 0.5687 & 0.7486 & 1.0000 & 0.7835 \\ 0.7581 & 0.8988 & 0.7124 & 0.5737 & 0.8790 & 0.7835 & 1.0000 \end{bmatrix}.$$

Definition 2. Let $(U, A, \tilde{R})$ be a fuzzy formal context. Given any $a_k, a_j \in A$, if there exists a path starting from a_k and ending at attribute a_j, and the correlation coefficient of attributes on this path is not less than δ, then the attribute similarity path of ith path is defined as follows:

$$PR_i^\delta = a_k \xrightarrow{r(a_k,a_2)} a_2 \cdots a_{t-1} \xrightarrow{r(a_{t-1},a_j)} a_j, \tag{7}$$

It should be pointed out that $r(a_k, a_2)$ on the arrow indicates the cost of getting from a_k to a_2 in the ith path. Then an attribute might appear in multiple paths due to the fact that the cost of going forward from attribute a_k may encounter multiple attributes not lower than δ. All attribute similarity paths are denoted as $PR^\delta = \{PR_1^\delta, PR_2^\delta, \ldots, PR_l^\delta\}$, where l represents the number of paths. In addition, the attributes within each attribute similarity path are clustered and denoted as attribute cluster, which is expressed as $C\mathcal{L} = \{CL_1, CL_2, \ldots, CL_l\}$ where $CL_i = \{a_k, a_2, \ldots, a_{t-1}, a_j | a_k \xrightarrow{r(a_k,a_2)} a_2 \cdots a_{t-1} \xrightarrow{r(a_{t-1},a_j)} a_j\}$.

Example 2. Assume $\delta = 0.7920$, then the attribute similarity paths also have $PR_1^\delta = a_1 \xrightarrow{1.00} a_1$, $PR_2^\delta = a_2 \xrightarrow{0.7943} a_5 \xrightarrow{0.8790} a_7$, $PR_3^\delta = a_3 \xrightarrow{0.8837} a_4$ and $PR_4^\delta = a_6 \xrightarrow{1.00} a_6$. In addition, suppose $\delta = 0.7750$, then the attribute similarity paths have $PR_1^\delta = a_1 \xrightarrow{1.00} a_1$, $PR_2^\delta = a_2 \xrightarrow{0.7943} a_5 \xrightarrow{0.8790} a_7$, $PR_3^\delta = a_2 \xrightarrow{0.7899} a_6 \xrightarrow{0.7835} a_7$ and $PR_4^\delta = a_3 \xrightarrow{0.8837} a_4$. From the above statement, it is evident that different δ can induce different attribute similarity paths. Next, the attribute clusters given by $\delta = 0.7920$ are $CL_1 = \{a_1\}$, $CL_2 = \{a_2, a_5, a_7\}$, $CL_3 = \{a_3, a_4\}$ and $CL_4 = \{a_6\}$. There exist intersecting attributes a_2 and a_7 in the attribute similarity paths PR_2^δ and PR_3^δ when $\delta = 0.7750$, and these paths are unioned. The final attribute clusters formed are $CL_1 = \{a_1\}$, $CL_2 = \{a_2, a_5, a_6, a_7\}$ and $CL_3 = \{a_3, a_4\}$.

Definition 3. For an attribute cluster CL_i, we define the representative attribute as follows:

$$Q(CL_i) = \{a_k \in A : \underset{a_j \in CL_i}{arg\,max}(rot_j)\}, \tag{8}$$

where $rot_j = \sum_{k \neq j, a_k \in CL_i} S(j,k)$ means the sum of correlation coefficient between attribute a_j with other attributes in the cluster CL_i. The representative attribute set is named $Q = \{Q(CL_1), Q(CL_2), \ldots, Q(CL_l)\}$.

Subsequently, we will continue to discuss the concept learning process in a new FFDC $(U, Q, \tilde{R}, D, J)$ formed by the representative attribute set Q. The process of extracting the representative attributes is shown in Algorithm 1.

Example 3 (*Continued with Example 1*)**.** Let $\delta = 0.7920$, for the attribute cluster CL_2, we obtain $rot_2 = S(2,5) + S(2,7) = 1.69$, $rot_5 = S(5,2) + S(5,7) = 1.67$ and $rot_7 = S(7,2) + S(7,5) = 1.78$. Then the attribute a_7 corresponding to the maximum value 1.78 is the representative attribute, which is $Q(CL_2) = \{a_7\}$. By a similar way, we obtain $Q = \{a_1, a_3, a_6, a_7\}$.

Then if $\delta = 0.7750$, for the attribute cluster CL_2, we have $rot_2 = 2.48$, $rot_5 = 2.42$, $rot_6 = 2.32$ and $rot_7 = 2.56$. Hence, $Q(CL_2) = \{a_7\}$. Finally, the representative attribute set is $Q = \{a_1, a_3, a_7\}$. Obviously, it is clearly known that different δ controls the size of the representative attribute set. That is to say, a larger value δ implies that fewer attributes satisfy the correlation coefficient, which leads to more attribute clusters and retains more representative attributes.

3.2. Fuzzy concept associative learning with distance metric learning

In this subsection, the similarities among different objects are measured by a distance metric between their attributes. A positive semi-definite matrix M is attained by optimizing the distance metrics among all objects. Thus, Eq. (2) can be used to describe a neighborhood similarity granule where $d_A(u, u_s) \in \mathbb{R}^+$, and satisfies non-negativity and symmetry. As is commonly known, associative memory [37] should

Algorithm 1: The representative attribute set extraction based on the correlation coefficient

Input: A fuzzy formal context $(U, A, \tilde{R})$ and threshold δ.
Output: The representative attribute set Q.

Initialize: $Q \leftarrow \emptyset$, $C \leftarrow \emptyset$;
for each $a_i \in A$ **do**
  for each $a_k \in A$ **do**
    Compute the correlation coefficient $r(a_i, a_k)$ from Definition 1;
    if $r(a_i, a_k) \geq \delta$ **then**
      $C(a_i) \leftarrow a_k$;
    end if
  end for
  $C \leftarrow C(a_i)$;
end for
for each $a_i \in A$ **do**
  for each $a_j \in A - \{a_i\}$ **do**
    if $C(a_i) \cap C(a_j) \neq \emptyset$ **then**
      $C(a_i) = C(a_i) \cup C(a_j)$ and $S \leftarrow \{a_j\}$; // S collects a set of other attributes whose intersection with a_i is non-empty.
    end if
  end for
  $CL_i \leftarrow C(a_i)$, $A = A - \{a_i\}$ and $A = A - S$;
  Compute rot_k in CL_i and select the representative attribute $a^* = \underset{a_k \in CL_i}{arg\max}(rot_k)$;
  $Q(CL_i) \leftarrow a^*$;
end for
$Q \leftarrow Q(CL_i)$.
Return Q.

be taken as the pivotal function of human brain computation. Furthermore, it is considered that the learning process of the human brain is a complicated one involving the generation, elimination, and modification of the relationships of neural information [38] in associative memory. Actually, clues are the key to associative learning, which are continuously associated with the knowledge in the brain during the learning process. This process might be performed repeatedly until all the knowledge corresponding to the clues is associated. Therefore, we introduce a novel fuzzy concept associative learning process with distance metric learning.

Similarly, in a new fuzzy formal context $(U, Q, \tilde{R})$, given $Y \subseteq U$ and $\tilde{C} \in \mathcal{F}(Q)$, a pair of cognitive operators $\tilde{T}_Q : \mathcal{P}(U) \to \mathcal{F}(Q)$ and $H_Q : \mathcal{F}(Q) \to \mathcal{P}(U)$ are given by:

$$\begin{aligned} &\tilde{T}_Q(Y)(a) = \bigwedge_{u \in Y} \tilde{R}(u, a), a \in Q, \\ &H_Q(\tilde{C}) = \left\{ u \in U : \forall a \in Q, \tilde{C}(a) \leq \tilde{R}(u, a) \right\}, \end{aligned} \tag{9}$$

In fact, the following discussion are based on $(U, Q, \tilde{R})$. Therefore, for the sake of convenience, we abbreviate two operators $\tilde{T}_Q$ and H_Q as $\tilde{T}$ and H, respectively.

Definition 4. Given a FFDC $(U, Q, \tilde{R}, D, J)$, in which $U/D = \{U^{d_1}, U^{d_2}, \ldots, U^{d_t}\}$. For $u \in U^{d_k}$, the neighborhood similarity granule of u is defined as follows:

$$N(u) = \{u_s \in U^{d_k} | d_Q(u, u_s) \leq \beta\}, \tag{10}$$

where β is a threshold.

Generally speaking, the smaller the distance $d_Q(u, u_s)$ between u and u_s is, the greater the similarity between them will be. And the threshold β is essential in concept learning process, which controls the size of $N(u)$. Then we denote all neighborhood similarity granules under U^{d_k} as $\mathcal{N}^{d_k} = \{N(u) | u \in U^{d_k}\}$. In fact, although $N(u)$ represents the set of some objects similar to u, there will be similar neighborhood similarity granules in $\mathcal{N}^{d_k}$ with inclusion relations. If the neighborhood similarity granules that satisfy the inclusion relationship are considered as clues of associative learning, which can reduce the time computational cost of learning concepts from multiple neighborhood similarity granules and is consistent with human thinking of learning concepts from the maximum clue. Therefore, the maximum clue in U^{d_k} is defined as

$$\begin{aligned} MC^{d_k}(u) = \{N(u) | u \in N(u) \wedge (\forall u \in N(u_i) \wedge N(u_i) \in \mathcal{N}^{d_k} \wedge N(u) \subseteq N(u_i) \\ \Rightarrow N(u_i) = N(u))\}. \end{aligned} \tag{11}$$

Note that the maximum clue of u includes all objects in U^{d_k} that is related to u, and $MC^{d_k}(u)$ may show a detailed and exhaustive description of u when we deliberate the issue of associative learning. The set of all maximum clues in U^{d_k} is denoted as $\mathcal{MC}^{d_k} = \{MC^{d_k}(u) | u \in U^{d_k}\}$. There might be many identical maximum clues in $\mathcal{MC}^{d_k}$, which shrinks the time consumption for associative learning from different maximum clues. For convenience, we abbreviate the maximum clue as clue $MC^{d_k}(u)$.

For $MC^{d_k}(u) \in \mathcal{MC}^{d_k}$, in the process of associative learning, clue $MC^{d_k}(u)$ is continuously associated with knowledge in the human brain, and finally concept is output according to the predetermined goals or personal preferences. From Eq. (9), it is obvious that $\left(H\tilde{T}(MC^{d_k}(u)), \tilde{T}(MC^{d_k}(u))\right)$ is a fuzzy concept. Now, we discuss the associative learning process from clue $MC^{d_k}(u)$.

For $MC^{d_k}(u) \in \mathcal{MC}^{d_k}$, the fuzzy concept associative subspace under U^{d_k} is represented as follows:

$$FAS^{d_k}(MC^{d_k}(u)) = \{(H\tilde{T}(X), \tilde{T}(X)) | X \subseteq MC^{d_k}(u)\}. \tag{12}$$

where $FAS^{d_k}(MC^{d_k}(u))$ means the set of fuzzy concepts corresponding to all subsets of clue $MC^{d_k}(u)$. However, it is an NP-hard problem to learn $FAS^{d_k}(MC^{d_k}(u))$ from $MC^{d_k}(u)$ when the clue is large enough. To address the aforementioned problem, we learn the efficient fuzzy concept associative subspace from $MC^{d_k}(u)$, that is,

$$\begin{aligned} EAS^{d_k}(MC^{d_k}(u)) = \{(H\tilde{T}(MC^{d_k}(u)), \tilde{T}(MC^{d_k}(u)))\} \cup \\ \{(H\tilde{T}(u_j), \tilde{T}(u_j)) | u_j \in MC^{d_k}(u)\}. \end{aligned} \tag{13}$$

In addition, the efficient fuzzy concept associative subspace about all clues $\mathcal{MC}^{d_k}$ in U^{d_k} is denoted as $EAS^{d_k} = \{EAS^{d_k}(MC^{d_k}(u_1)), EAS^{d_k}(MC^{d_k}(u_2)), \ldots, EAS^{d_k}(MC^{d_k}(u_r))\}$, where r means the number of clues in $\mathcal{MC}^{d_k}$.

As mentioned before, $EAS^{d_k}(MC^{d_k}(u))$ is a set of fuzzy concepts related to $MC^{d_k}(u)$. However, we are more concerned about the fuzzy concept that emerges after the process of association ends. For a fuzzy set $\tilde{B}$ and $a \in Q$, the dominance operator $\diamond$ about a is denoted as $\tilde{B}^{\diamond}(a) = \{u_j \in U | \tilde{R}(u_j, a) \geq \tilde{B}(a)\}$. Obviously, $\tilde{B}^{\diamond}(a)$ is a collection of objects whose fuzzy membership values with respect to attribute a are greater than $\tilde{B}(a)$.

Definition 5. For $(H\tilde{T}(X), \tilde{T}(X)) \in EAS^{d_k}(MC^{d_k}(u))$, then the extent-intent validity is given by:

$$Eiv(H\tilde{T}(X), \tilde{T}(X)) = \sum_{a \in Q} \frac{|H\tilde{T}(X) \cap \tilde{T}(X)^{\diamond}(a)|}{|\tilde{T}(X)^{\diamond}(a)|}, \tag{14}$$

where $|*|$ means the number of $*$.

More precisely, $Eiv(H\tilde{T}(X), \tilde{T}(X))$ means the effective contribution of extent $H\tilde{T}(X)$ and intent $\tilde{T}(X)$ to the formation of fuzzy concept $(H\tilde{T}(X), \tilde{T}(X))$. The larger $Eiv(H\tilde{T}(X), \tilde{T}(X))$ is, the more relevant the fuzzy concept outputted and clue through associative learning is.

Definition 6. For $EAS^{d_k}(MC^{d_k}(u)) \in EAS^{d_k}$, then the valid fuzzy concept associative subspace is given by:

$$\begin{aligned} VCAS^{d_k} = \{(H\tilde{T}(Y), \tilde{T}(Y)) | \underset{(H\tilde{T}(X), \tilde{T}(X)) \in EAS^{d_k}(MC^{d_k}(u))}{arg\max} Eiv(H\tilde{T}(X), \tilde{T}(X)), \\ \forall EAS^{d_k}(MC^{d_k}(u)) \in EAS^{d_k}\}. \end{aligned} \tag{15}$$

Definition 6 explains that a most valuable fuzzy concept is learned and derived from the given fuzzy concept associative subspace $EAS^{d_k}(MC^{d_k}(u))$. Finally, the valid fuzzy concept associative subspace under each decision class is integrated into the overall concept space, which is called $VCAS^d = \{VCAS^{d_1}, VCAS^{d_2}, \ldots, VCAS^{d_t}\}$. Next, Algorithm 2 illustrates the process of constructing a valid fuzzy concept associative space.

Algorithm 2: Constructing valid fuzzy concept associative space

Input: A FFDC $(U, Q, \tilde{R}, D, J)$ and threshold β.
Output: Valid fuzzy concept associative space $VCAS^d$.

Initialize: $VCAS^d \leftarrow \emptyset$;
for each $U^{d_k} \in U/D$ **do**
  Initialize: $\mathcal{N}^{d_k} \leftarrow \emptyset$, $\mathcal{MC}^{d_k} \leftarrow \emptyset$ and $VCAS^{d_k} \leftarrow \emptyset$;
  for each $u_j \in U^{d_k}$ **do**
    Compute the neighborhood similarity granule $N(u_j)$ from Definition 4 and $\mathcal{N}^{d_k} \leftarrow N(u_j)$;
  end for
  while $\mathcal{N}^{d_k} \neq \emptyset$ **do**
    Sort the neighborhood similarity granule in $\mathcal{N}^{d_k}$ in descending order of their number;
    for each $u_j \in U^{d_k}$ **do**
      for $u_i \in U^{d_k} - \{u_j\}$ **do**
        if $N(u_i) \subseteq N(u_j)$ **then**
          $MC^{d_k}(u_j) \leftarrow N(u_j)$ according to Eq. (11) and $\mathcal{N}^{d_k} = \mathcal{N}^{d_k} - N(u_i)$;
        else
          continue
        end if
      end for
      $\mathcal{MC}^{d_k} \leftarrow MC^{d_k}(u_j)$;
    end for
  end while
  for $MC^{d_k}(u_j) \in \mathcal{MC}^{d_k}$ **do**
    Compute $(H\tilde{T}(MC^{d_k}(u_j)), \tilde{T}(MC^{d_k}(u_j)))$ and $Eiv(H\tilde{T}(MC^{d_k}(u_j)), \tilde{T}(MC^{d_k}(u_j)))$;
    $EAS^{d_k}(MC^{d_k}(u)) \longleftarrow (H\tilde{T}(MC^{d_k}(u_j)), \tilde{T}(MC^{d_k}(u_j)))$;
    for $u_s \in MC^{d_k}(u_j)$ **do**
      Construct a fuzzy concept $(H\tilde{T}(u_s), \tilde{T}(u_s))$ and calculate $Eiv(H\tilde{T}(u_s), \tilde{T}(u_s))$;
      $EAS^{d_k}(MC^{d_k}(u)) \longleftarrow (H\tilde{T}(u_s), \tilde{T}(u_s))$;
    end for
    Select a fuzzy concept $(H\tilde{T}(Y), \tilde{T}(Y))$ based on $arg\max Eiv(H\tilde{T}(Y), \tilde{T}(Y))$ where $(H\tilde{T}(Y), \tilde{T}(Y)) \in EAS^{d_k}(MC^{d_k}(u))$;
    $VCAS^{d_k} \leftarrow (H\tilde{T}(Y), \tilde{T}(Y))$;
  end for
  $VCAS^d \leftarrow VCAS^{d_k}$.
end for
Return $VCAS^d$.

Example 4 (*Proceeded with Example 1*)**.** Suppose $\delta = 0.7920$, $\beta = 0.4$, and the nearest neighbors $k = 2$. We can obtain M for four attributes a_1, a_3, a_6, a_7 is 1.43, 1.13, 1.23 and 0.59.

For the objects in decision class U^{d_1}, we can calculate the neighborhood similarity granule $N(u_1) = N(u_6) = \{u_1, u_6\}$, $N(u_2) = \{u_2\}$, $N(u_3) = N(u_4) = N(u_5) = \{u_3, u_4, u_5\}$ and $N(u_7) = \{u_7\}$. Then the clues are $MC^{d_1}(u_1) = \{u_1, u_6\}$, $MC^{d_1}(u_2) = \{u_2\}$, $MC^{d_1}(u_3) = \{u_3, u_4, u_5\}$ and $MC^{d_1}(u_7) = \{u_7\}$. For $MC^{d_1}(u_1)$ and $MC^{d_1}(u_3)$, the fuzzy concept associative subspaces and the extent-intent validities are represented in Table 3. Thus, the fuzzy concepts $\left(\{u_6\}, \{\frac{a_1}{0.69}, \frac{a_3}{0.84}, \frac{a_6}{0.57}, \frac{a_7}{0.88}\}\right)$ and $\left(\{u_1, u_3, u_4, u_5, u_6\}, \{\frac{a_1}{0.32}, \frac{a_3}{0.09}, \frac{a_6}{0.30}, \frac{a_7}{0.53}\}\right)$ corresponding to the maximum value 2.83 and 3.38 are the most valuable fuzzy concepts output through the process of associative learning. Next, we can obtain the valid fuzzy concepts associative subspace $VCAS^{d_1}$ in Table 4.

In addition, for the objects in decision class U^{d_2}, the neighborhood similarity granules are $N(u_8) = \{u_8, u_{13}\}$, $N(u_9) = \{u_9\}$, $N(u_{10}) = N(u_{11}) = \{u_{10}, u_{11}\}$, $N(u_{12}) = \{u_{12}, u_{13}\}$ and $N(u_{13}) = \{u_8, u_{12}, u_{13}\}$. Then we have $MC^{d_2}(u_9) = \{u_9\}$, $MC^{d_2}(u_{10}) = \{u_{10}, u_{11}\}$ and $MC^{d_2}(u_{13}) = \{u_8, u_{12}, u_{13}\}$. Employing a similar way, the valid fuzzy concept associative subspace $VCAS^{d_2}$ is shown in Table 4.

3.3. Concept clustering

In the above section, constructing a valid fuzzy concept associative space based on maximum clues have been explored. But in practice, there exist a lot of repetitive and interfering information between fuzzy concepts, which will reduce the computational efficiency of model. Numerous fuzzy concepts provide different significance and merit in concept associative learning. Therefore, only the more important fuzzy concepts can be retained and compressed because human memory is limited. To address appropriate fuzzy ontologies for concept cognition, many concept fusion mechanisms have been studied, such as progressive fuzzy three-way concept [9], fuzzy concept clustering [10,24,27] and interval-intent fuzzy concept clustering [28]. These method of generating pseudo-concepts are fused by fixed weights, which are not characterized by the information of concepts, which may lead to the transfer of conceptual preferences in the process of fusing.

As mentioned above, we note that M is a diagonal matrix in Eq. (5), which corresponds to learning a metric where different axes is assigned different weights. More precisely, weights are used to evaluate the importance of attributes. We denote $diag(M) = (\omega(a_1), \omega(a_2), \ldots, \omega(a_q))$ for convenience. In this subsection, we study an innovative fuzzy concept clustering approach to compress the valid fuzzy concept associative space.

Definition 7. For $C_r^{d_k} \subseteq VCAS^{d_k}$ and $(Y_1, \tilde{B}_1), (Y_2, \tilde{B}_2), \ldots, (Y_s, \tilde{B}_s) \in C_r^{d_k}$, if there exists $Y_1 \subseteq Y_2 \subseteq \cdots \subseteq Y_s$, then $(Y_s, \tilde{B}_s)$ is a supremum fuzzy concept, the extent and intent of pseudo-concept $(\mathbf{Y}_r, \tilde{\mathbf{B}}_r)$ is denoted as:

$$\begin{aligned} \mathbf{Y}_r &= Y_1 \cup Y_2 \cup \cdots \cup Y_s, \\ \tilde{\mathbf{B}}_r &= \frac{\sum_{i=1}^{s} \tilde{B}_i \cdot diag(M)}{s}. \end{aligned} \tag{16}$$

in which s means the cardinality of fuzzy concepts.

In fact, we can see that the intent of pseudo-concept is the average information after assigning weights to the intents of concepts in $C_r^{d_k}$. Meanwhile, pseudo-concept is integrated by fuzzy concepts that have extents with inclusion relation in $C_r^{d_k}$, which reduces space storage and eliminate cognitive limitations. Pseudo-concept can be seen as a reintegrated representation of $C_r^{d_k}$. Algorithm 3 outlines the clustering process of valid fuzzy concept associative space.

Example 5. Proceeded with Example 4. Table 5 shows the pseudo-concept space $\mathfrak{P}$ from Definition 7. There are 2 and 2 pseudo-concepts in $\mathfrak{P}^{d_1}$ and $\mathfrak{P}^{d_2}$, respectively.

3.4. Class prediction

It should be noticed that the class prediction of testing sample is mainly measured by the Euclidean distance between testing sample and existing fuzzy concept clustering space [28–30]. Generally speaking, the similarities between sample and pseudo-concepts is only determined by straight-line distance, ignoring the correlation between attributes and decision classes of the existing samples. In such case, it is necessary to propose a new method to measure similarity.

Definition 8. Given a testing sample x_r, $(x_r, \tilde{T}(x_r))$ is a new fuzzy concept, then the similarity between $\tilde{T}(x_r)$ and pseudo-concept $(\mathbf{Y}_i, \tilde{\mathbf{B}}_i)$ is termed as:

$$Sim(\tilde{T}(x_r), \tilde{\mathbf{B}}_i) = \sqrt{(\tilde{T}(x_r) - \tilde{\mathbf{B}}_i)^T M (\tilde{T}(x_r) - \tilde{\mathbf{B}}_i)}, \tag{17}$$

where M is a diagonal matrix in Eq. (5).

Table 3
Fuzzy concept associative subspace and extent-intent validity.

Clue	Fuzzy concept associative subspaces	Validity
$MC^{d_1}(u_1)$	$(\{u_6\}, \{\frac{a_1}{0.69}, \frac{a_3}{0.84}, \frac{a_6}{0.57}, \frac{a_7}{0.88}\})$	2.83
	$(\{u_1\}, \{\frac{a_1}{0.73}, \frac{a_3}{0.59}, \frac{a_6}{0.40}, \frac{a_7}{0.65}\})$	1.33
	$(\{u_1, u_6\}, \{\frac{a_1}{0.69}, \frac{a_3}{0.59}, \frac{a_6}{0.40}, \frac{a_7}{0.65}\})$	2.33
$MC^{d_1}(u_3)$	$(\{u_3, u_6\}, \{\frac{a_1}{0.44}, \frac{a_3}{0.09}, \frac{a_6}{0.46}, \frac{a_7}{0.67}\})$	2.45
	$(\{u_4, u_6\}, \{\frac{a_1}{0.37}, \frac{a_3}{0.33}, \frac{a_6}{0.30}, \frac{a_7}{0.73}\})$	2.13
	$(\{u_1, u_5, u_6\}, \{\frac{a_1}{0.32}, \frac{a_3}{0.28}, \frac{a_6}{0.31}, \frac{a_7}{0.53}\})$	2.2
	$(\{u_1, u_3, u_4, u_5, u_6\}, \{\frac{a_1}{0.32}, \frac{a_3}{0.09}, \frac{a_6}{0.30}, \frac{a_7}{0.53}\})$	3.38

Table 4
Valid fuzzy concept associative space $VCAS^d$ in Table 2.

Symbol	Fuzzy concepts learned through associative learning
$VCAS^{d_1}$	$(\{u_6\}, \{\frac{a_1}{0.69}, \frac{a_3}{0.84}, \frac{a_6}{0.57}, \frac{a_7}{0.88}\})$
	$(\{u_1, u_2, u_4, u_6, u_7\}, \{\frac{a_1}{0.10}, \frac{a_3}{0.33}, \frac{a_6}{0.04}, \frac{a_7}{0.10}\})$
	$(\{u_1, u_3, u_4, u_5, u_6\}, \{\frac{a_1}{0.32}, \frac{a_3}{0.09}, \frac{a_6}{0.30}, \frac{a_7}{0.53}\})$
	$(\{u_7\}, \{\frac{a_1}{0.96}, \frac{a_3}{0.88}, \frac{a_6}{0.45}, \frac{a_7}{0.39}\})$
$VCAS^{d_2}$	$(\{u_8, u_{10}, u_{11}, u_{12}, u_{13}\}, \{\frac{a_1}{0.07}, \frac{a_3}{0.31}, \frac{a_6}{0.09}, \frac{a_7}{0.27}\})$
	$(\{u_9, u_{11}\}, \{\frac{a_1}{0.61}, \frac{a_3}{0.40}, \frac{a_6}{0.20}, \frac{a_7}{0.02}\})$
	$(\{u_{11}\}, \{\frac{a_1}{0.67}, \frac{a_3}{0.67}, \frac{a_6}{0.88}, \frac{a_7}{0.79}\})$

Algorithm 3: Cognitive process of valid fuzzy concept associative clustering space

Input: A valid fuzzy concept associative space $VCAS^d$.
Output: The pseudo-concept space $\mathfrak{P} = \{\mathfrak{P}^{d_1}, \mathfrak{P}^{d_2}, \cdots, \mathfrak{P}^{d_t}\}$.

for each $VCAS^{d_k} \in VCAS^d$ **do**
  Sort the elements in $VCAS^{d_k}$ in descending order of the number of extents;
  for each $(Y_i, \tilde{B}_i) \in VCAS^{d_k}$ **do**
    for $(Y_j, \tilde{B}_j) \in VCAS^{d_k}$ **do**
      if $Y_j \subseteq Y_i$ **then**
        $C_i^{d_k} \leftarrow (Y_j, \tilde{B}_j)$;
      else
        continue
      end if
    end for
    Cluster $C_i^{d_k}$ to generate a pseudo-concept $(\mathbf{Y}_i, \tilde{\mathbf{B}}_i)$;
    $\mathfrak{P}^{d_k} \leftarrow (\mathbf{Y}_i, \tilde{\mathbf{B}}_i)$;
    $VCAS^{d_k} = VCAS^{d_k} - C_i^{d_k}$;
  end for
  $\mathfrak{P} \leftarrow \mathfrak{P}^{d_k}$.
end for
Return $\mathfrak{P}$.

Algorithm 4: Class prediction of testing sample

Input: The pseudo-concept space $\mathfrak{P}$ and the fuzzy membership degree $\tilde{T}$ of testing sample x_r.
Output: Class prediction L_r of testing sample x_r.

for $\mathfrak{P}^{d_k} \in \mathfrak{P}$ **do**
  for $(\mathbf{Y}_i, \tilde{\mathbf{B}}_i) \in \mathfrak{P}^{d_k}$ **do**
    Compute $Sim(\tilde{T}(x_r), \tilde{\mathbf{B}}_i)$ according to Definition 8 and $l_r \leftarrow Sim(\tilde{T}(x_r), \tilde{\mathbf{B}}_i)$;
  end for
end for
Have $L_r \leftarrow arg\min(min(l_r))$.
Return L_r.

As is well known, $Sim(\tilde{T}(x_r), \tilde{\mathbf{B}}_i)$ is a distance similarity. The smaller the distance $Sim(\tilde{T}(x_r), \tilde{\mathbf{B}}_i)$ is, the stronger the relationship between two fuzzy concepts is. Therefore, we can predict the class of new sample by obtaining the similarity between new sample and pseudo-concepts in $\mathfrak{P}$.

3.5. Overall procedure and complexity analysis

As summarized above, Fig. 2 describes the comprehensive flowchart of FCADML, which includes three parts, namely (1) Constructing a new fuzzy formal decision context; (2) Constructing valid fuzzy concept associative space; (3) Class prediction. Given a fuzzy formal context with fourteen attributes and two decision classes, we will explain the idea of flowchart. Due to the high correlation between some attributes, which increases the time consumption of computing fuzzy concepts, highly similar attributes are clustered into an attribute cluster from Definition 2. Then a representative attribute set from all attribute clusters is selected to ultimately construct a new FFDC with four conditional attributes. Subsequently, we propose the neighborhood similarity granules controlled by threshold with distance metric learning and maximum clues for associative learning. It should be noted that $N(1)$ and $MC(1)$ represent the neighborhood similarity granule and the maximum cue of object u_1, respectively, and other symbols also have the same meaning. In the process of associative learning, maximum clues are constantly associated with the knowledge in the human brain, and valid fuzzy concept space $VCAS^d$ is finally output based on extent-intent validity. Pseudo-concept space is discussed from Definition 7. At last, finding the pseudo-concept that is closest to the testing sample to determine the decision class.

In addition, suppose t is the cardinality of decision class. The time complexity of Algorithm 1 for extracting representative attribute set through the correlation coefficient matrix does not exceed $O(|U\|A|^2 + |A\|A-1|)$. Next, in Algorithm 2, constructing valid fuzzy concept associative space includes three steps: (1) Obtaining the neighborhood similarity granules in Steps 3–6 which can be taken in $O(|U|^2)$; (2) Finding maximum clues in Steps 8–20 which can be measured within $O(|\mathcal{N}^{d_k} \| U|^2)$ where $\mathcal{N}^{d_k}$ means the set of neighborhood similarity granules; (3) Computing fuzzy concept associative space and valid fuzzy concept associative space in Steps 21–29 which the time complexity is $O(|\mathcal{MC}^{d_k} \| U|^2|A|)$ where $\mathcal{MC}^{d_k}$ is the maximum clue set in decision class d_k. All in all, the time complexity of running Algorithm 2 is $O(t|\mathcal{MC}^{d_k} \| U|^2|A|)$ in which t is the cardinality of decision class.

Table 5
Pseudo-concept space $\mathfrak{P}$.

Symbol	Pseudo-concepts
$\mathfrak{P}^{d_1}$	$(\{u_1,u_3,u_4,u_5,u_6\},\{\frac{a_1}{0.82},\frac{a_3}{0.55},\frac{a_6}{0.59},\frac{a_7}{0.14}\})$ $(\{u_1,u_2,u_4,u_6,u_7\},\{\frac{a_1}{0.86},\frac{a_2}{0.71},\frac{a_6}{0.33},\frac{a_7}{0.05}\})$
$\mathfrak{P}^{d_2}$	$(\{u_8,u_{10},u_{11},u_{12},u_{13}\},\{\frac{a_1}{0.60},\frac{a_3}{0.57},\frac{a_6}{0.66},\frac{a_7}{0.11}\})$ $(\{u_9,u_{11}\},\{\frac{a_1}{0.99},\frac{a_3}{0.47},\frac{a_6}{0.27},\frac{a_7}{0.01}\})$

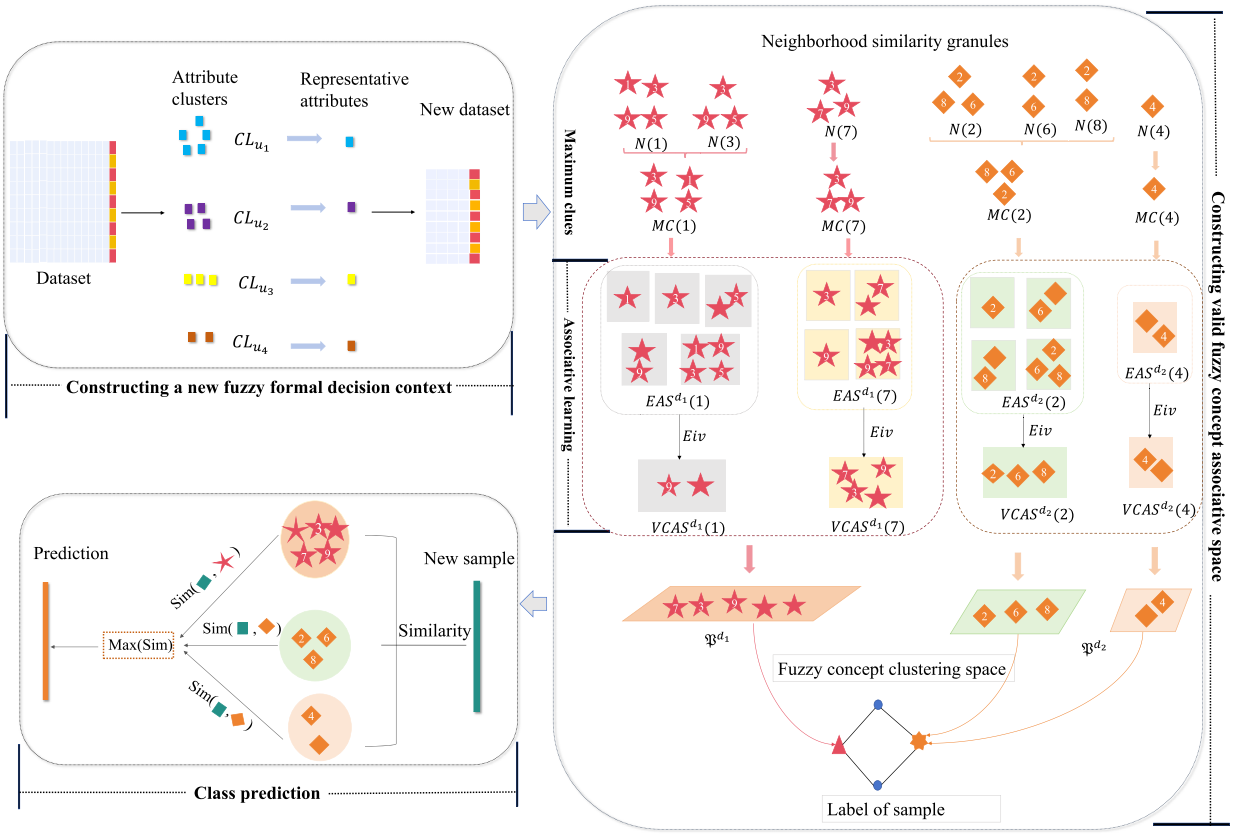

Fig. 2. The overall procedure of the proposed method.

In the clustering process, the fuzzy concepts that have already been clustered will not appear in the subsequent clustering process, so the number of fuzzy concepts to consider will gradually decrease each time. Hence, the time complexity of Algorithm 3 in the worst-case scenario does not exceed $O(|U|^2|A|)$. With respect to Algorithm 4, the running time for class prediction of testing sample is $O(|\mathfrak{P}\parallel A|)$. Hence, the total time complexity of the proposed model FCADML is $O(max\{|U\parallel A|^2+|A\|A-1|,t|\mathcal{MC}^{d_k}\|U|^2|A|\})$.

4. Experimental analyses

To demonstrate the classification performance of model FCADML, we will make a comparison between it and concept-cognitive learning as well as fuzzy classification algorithms. Meanwhile, we also compare it with machine learning classification algorithms. We select totally thirteen datasets from UCI[1] and gene[2] data set repositories, which show in Table 6.

4.1. Experimental setting

Before the experiment starts, the values of all attributes are normalized by Max–Min normalization preprocessing to ensure the fuzzy environment of dataset, specifically given by:

$$\tilde{R}(u_i,a_j)=\frac{f(u_i,a_j)-\min(f(a_j))}{\max(f(a_j))-\min(f(a_j))},$$

where $f(u_i,a_j)$ is the initial value of object u_i to attribute a_j. Additionally, $\max(f(a_j))$ and $\min(f(a_j))$ are maximum and minimum values of attribute a_j about all objects, respectively.

In the experiment, we compare FCADML with three concept-cognitive learning classification algorithms, namely, DMPWFC [10], ILMPFTC [9], FCLM [24] and four fuzzy classification algorithms which are PIFWKNN [39], BM-FKNN [40], S3OFIS [41] and FSMBGD [42]. Furthermore, we also compare FCADML with seven machine learning classification algorithms [43–48], including Complex Tree (CT), Classification and Regression Tree (CART), Root-Sum-Square (RSS), Decision Tree (DT), Boosting, K-Nearest Neighbor (KNN) with $k=3$, and Gaussian Kernel Function SVM (GSVM). Notice that parameters δ and β play an important role in constructing attribute clusters and neighborhood similarity granules. Therefore, in the present experiment, parameters δ

[1] Dataset source: http://archive.ics.uci.edu/.
[2] Dataset source: https://jundongl.github.io/scikit-feature/datasets.html.

Table 6
Data description.

ID	Dataset	Object	Attribute	Class
1	Connectionist	208	60	2
2	Libras	360	90	15
3	Wdbc	569	30	2
4	Appendicitis	106	7	2
5	Air	359	64	3
6	Derm	366	34	6
7	Phoneme	5404	5	2
8	Waveform	5000	21	3
9	Lung_disarete	73	325	7
10	Colonstd	62	2000	2
11	Yale	165	1024	15
12	WarpAR10P	130	2400	10
13	Lung	203	3312	5

Table 7
Accuracy comparison (mean ± standard deviation%) under FCADML and other algorithms.

ID	(δ, β)	FCADML	DMPWFC	ILMPFTC	FCLM	PIFWKNN	BM-FKNN	S3OFIS	FSMBGD
1	(0.85,0.5)	85.62 ± 8.39	88.82 ± 6.50	**91.01 ± 5.65**	83.98 ± 12.17	87.49 ± 3.15	60.00 ± 3.12	83.66 ± 6.28	79.06 ± 5.64
2	(1.00,0.10)	**85.00 ± 5.41**	60.01 ± 6.54	84.81 ± 4.24	65.03 ± 3.58	66.11 ± 4.00	61.56 ± 5.13	73.33 ± 7.24	68.31 ± 9.86
3	(1.00,0.20)	**95.96 ± 2.28**	86.51 ± 3.70	91.67 ± 2.72	80.79 ± 5.29	95.78 ± 1.40	95.96 ± 1.00	94.90 ± 0.74	95.48 ± 5.13
4	(1.00,0.35)	**85.84 ± 3.38**	68.15 ± 13.31	71.38 ± 16.71	50.45 ± 25.41	85.02 ± 8.02	19.83 ± 8.55	8.48 ± 2.09	78.97 ± 4.84
5	(0.90,0.30)	**96.67 ± 2.11**	47.18 ± 5.00	95.63 ± 4.20	54.11 ± 5.67	87.48 ± 3.51	86.08 ± 4.68	86.07 ± 2.15	92.56 ± 5.86
6	(0.90,0.50)	**95.09 ± 3.90**	87.58 ± 3.88	94.93 ± 3.21	37.62 ± 5.00	94.42 ± 4.35	84.69 ± 6.09	94.80 ± 2.64	95.04 ± 3.35
7	(0.85,0.05)	76.41 ± 0.34	75.12 ± 1.45	67.21 ± 1.40	72.15 ± 1.41	**85.64 ± 1.05**	29.35 ± 1.59	5.22 ± 0.49	73.69 ± 2.15
8	(0.80,0.45)	80.12 ± 1.09	56.17 ± 1.29	76.35 ± 1.28	68.25 ± 1.06	82.34 ± 2.12	59.94 ± 1.52	6.64 ± 1.03	**84.52 ± 4.65**
9	(0.85,0.05)	70.00 ± 10.78	**87.80 ± 8.05**	87.80 ± 8.05	87.17 ± 7.94	80.67 ± 11.66	2.58 ± 0.51	82.00 ± 6.87	61.25 ± 5.74
10	(0.95,0.05)	**76.03 ± 13.75**	65.34 ± 19.55	65.34 ± 19.55	52.35 ± 19.41	69.87 ± 15.02	0.52 ± 0.04	74.10 ± 10.83	70.74 ± 2.66
11	(1.00,0.05)	55.76 ± 5.50	66.54 ± 8.60	66.54 ± 8.60	54.51 ± 10.64	64.24 ± 4.49	2.03 ± 0.22	66.67 ± 7.42	**68.69 ± 9.30**
12	(1.00,0.05)	**80.00 ± 6.88**	47.49 ± 7.53	47.49 ± 7.53	43.66 ± 7.81	46.92 ± 16.85	0.68 ± 0.15	47.69 ± 4.39	50.38 ± 5.34
13	(0.90,0.05)	94.06 ± 3.82	94.93 ± 3.66	94.93 ± 3.66	92.17 ± 12.04	**97.55 ± 2.99**	1.18 ± 0.05	94.55 ± 5.70	84.79 ± 5.98
Ave.± SD		**82.81 ± 5.20**	71.66 ± 6.85	79.62 ± 6.68	64.79 ± 9.44	80.27 ± 6.05	38.80 ± 2.51	62.93 ± 4.45	77.19 ± 5.42
Rank		**2.58**	4.88	3.50	6.15	3.54	6.19	4.85	3.69
Win/tie/loss		**7/0/6**	1/0/12	1/0/12	0/0/13	2/0/11	0/0/13	0/0/13	2/0/11

and β are set from 0.55 to 1.0, and 0.05 to 0.50, respectively, with a step size of 0.05 to predominate attribute clusters and neighborhood similarity granules. To maintain the fairness of the experiment, we employ five-fold cross-validation in our experiment. In other words, each dataset is randomly divided into five segments, where four segments are combined as a training set and the other segment is utilized as a testing set. Intuitively, 80% of each dataset is used for training and the remaining 20% is used for testing. Then the final output is the average value of five testing results to estimate the classification performance. All experiments are conducted in MATLAB 2021a on a personal computer which is furnished with Intel(R) Core(TM) i7-4790 CPU @ 3.6 GHz and 16 GB memory.

4.2. Results and analyses

The results of accuracy and the optimal parameters δ and β of the proposed algorithm FCADML and seven classification algorithms on thirteen datasets are displayed in Table 7. The last row presents the average accuracy and standard deviation, with the underlined bold highlighting the superior accuracy performance in contrast to other algorithms. As seen in this table, we can demonstrate that PIFWKNN and FSMBGD achieve the best accuracy twice respectively, while FCADML outperforms the best performance on seven datasets. Meanwhile, FCADML has higher average accuracy and lower standard deviation than other algorithms, which shows that the performance of FCADML outperforms other seven algorithms. The average accuracy of FCADML is increased by 4.01% and 3.16% when compared to ILMPFTC and PIFWKNN on all selected datasets. In short, the above results indicate that FCADML is superior to the other models in classification tasks.

In addition, to further evaluate the performance of FCADML, Table 8 records the comparison of classification accuracy with seven machine learning algorithms. The results show that FCADML achieves higher accuracy of 7 times, while CART, KNN and GSVM realize a maximum value on 1, 3 and 2 out of thirteen datasets, respectively. Then FCADML improves the average accuracy by 1.37% when compared to KNN. Figs. 3–4 intuitively depict the accuracy comparison between CCL, fuzzy classification and machine learning algorithms. As presented from two figures, the accuracy of FCADML has smaller fluctuation range than other fourteen algorithms in most datasets.

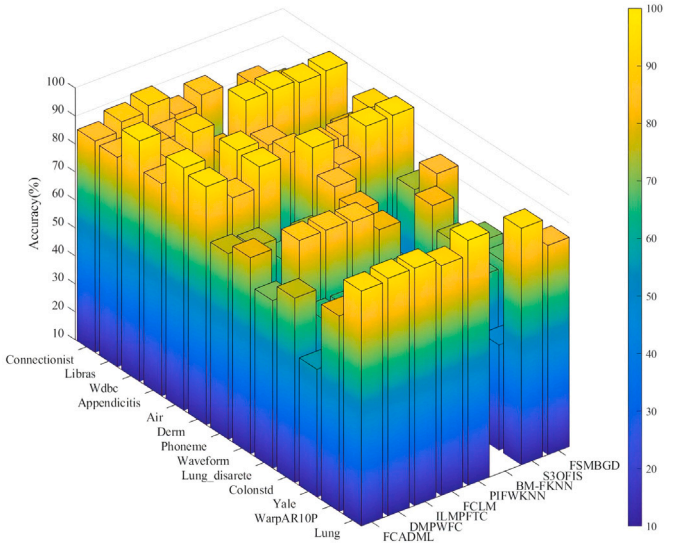

Fig. 3. Accuracy comparison with CCL and fuzzy classification algorithms.

Table 8
Accuracy comparison (mean ± standard deviation%) under FCADML and seven machine learning algorithms.

ID	(δ, β)	FCADML	CT	CART	RSS	DT	Boosting	KNN	GSVM
1	(0.85,0.50)	**85.62 ± 8.39**	71.63 ± 9.34	77.92 ± 4.39	62.04 ± 14.73	63.43 ± 7.82	75.02 ± 3.46	83.64 ± 4.05	81.71 ± 5.05
2	(1.00,0.10)	**85.00 ± 5.41**	62.78 ± 8.86	62.22 ± 4.75	36.94 ± 5.43	19.44 ± 6.73	11.67 ± 4.12	77.50 ± 3.85	81.39 ± 7.19
3	(1.00,0.20)	95.96 ± 2.28	91.57 ± 1.58	91.38 ± 2.90	86.64 ± 2.28	88.22 ± 4.15	94.38 ± 2.37	95.84 ± 1.00	**97.36 ± 1.77**
4	(1.00,0.35)	**85.84 ± 3.38**	83.90 ± 8.74	80.22 ± 3.75	83.03 ± 8.61	69.65 ± 16.86	82.08 ± 6.19	84.89 ± 4.03	85.71 ± 10.10
5	(0.90,0.30)	**96.67 ± 2.11**	81.62 ± 4.17	82.45 ± 2.69	45.40 ± 3.93	47.63 ± 2.93	57.93 ± 3.08	96.37 ± 1.88	91.65 ± 4.04
6	(0.90,0.50)	95.09 ± 3.91	95.08 ± 1.22	93.73 ± 4.12	55.43 ± 9.96	62.27 ± 7.89	78.97 ± 7.27	**96.99 ± 2.45**	95.07 ± 1.85
7	(0.85,0.05)	76.41 ± 0.34	85.38 ± 0.71	85.92 ± 1.19	70.65 ± 1.15	78.09 ± 1.47	77.72 ± 0.90	**88.47 ± 1.16**	80.50 ± 1.29
8	(0.80,0.45)	80.12 ± 1.09	77.32 ± 1.27	75.54 ± 0.66	77.80 ± 2.32	67.74 ± 1.29	69.44 ± 1.59	80.48 ± 1.08	**86.12 ± 0.72**
9	(0.85,0.05)	70.00 ± 10.78	45.24 ± 22.88	52.95 ± 18.89	49.71 ± 14.11	6.95 ± 6.91	42.48 ± 14.09	**84.86 ± 8.90**	28.67 ± 14.92
10	(0.95,0.05)	76.03 ± 13.75	77.31 ± 10.55	**80.77 ± 8.75**	64.23 ± 16.45	74.10 ± 9.08	77.56 ± 5.98	72.31 ± 8.19	64.49 ± 12.59
11	(1.00,0.05)	**55.76 ± 5.50**	42.42 ± 5.67	51.51 ± 6.78	33.33 ± 11.54	6.06 ± 10.50	9.70 ± 6.91	52.73 ± 6.28	0.61 ± 1.36
12	(1.00,0.05)	**80.00 ± 6.88**	66.15 ± 3.22	68.46 ± 9.58	32.31 ± 12.64	25.38 ± 9.65	32.31 ± 5.83	53.85 ± 7.20	1.54 ± 2.11
13	(0.90,0.05)	**94.06 ± 3.82**	88.17 ± 8.55	83.79 ± 4.92	68.46 ± 5.67	85.28 ± 8.55	78.80 ± 4.52	94.06 ± 3.78	68.52 ± 5.24
Ave.± SD		**82.81 ± 5.20**	74.51 ± 6.67	75.91 ± 5.64	58.92 ± 8.37	53.40 ± 7.22	60.62 ± 5.10	81.69 ± 4.14	66.41 ± 5.25
Rank		**2.12**	4.08	4.08	6.54	6.69	5.69	2.42	4.38
Win/tie/loss		**7/0/6**	0/0/13	1/0/12	0/0/13	0/0/13	0/0/13	3/0/10	2/0/11

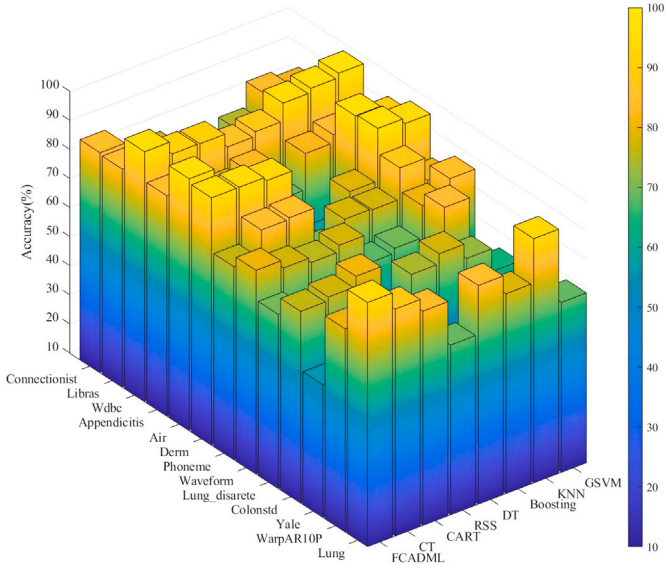

Fig. 4. Accuracy comparison with machine learning classification algorithms.

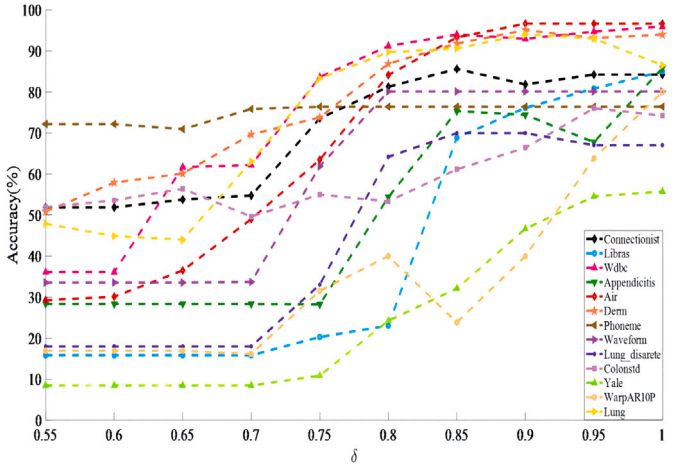

Fig. 5. Accuracy comparisons various with δ on thirteen datasets.

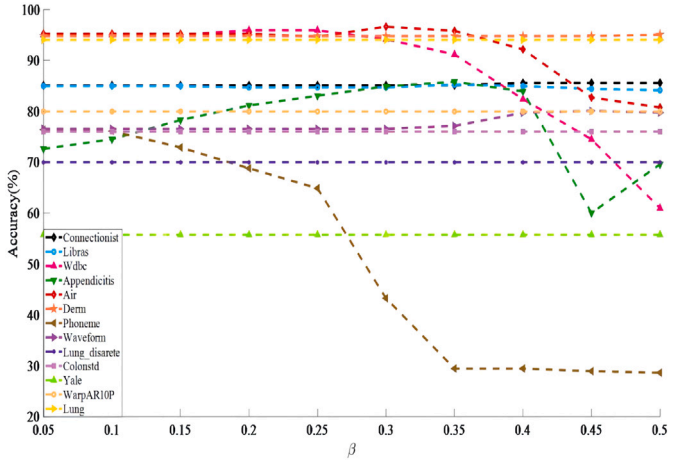

Fig. 6. Accuracy comparisons various with β on thirteen datasets.

4.3. Parametric analyses

The algorithm FCADML involves two parameters, where parameter δ reflects the size of the attribute cluster, which further determines the importance of the representative attributes, and parameter β plays a critical role in constructing the neighborhood similarity granule, which measures the amount of information and classification performance. Figs. 5–7 record the accuracies of FCADML in different parameters. Fig. 5 depicts the classification accuracy as δ changes, from which we see that the accuracy have been fluctuating and rising significantly as the parameter changes with the highest fluctuation at [0.7,0.85]. Meanwhile, most datasets achieve optimal accuracy in the range of [0.9,1.0], indicating that larger parameters lead to fewer attribute correlations being satisfied, resulting in more attribute clusters and thus extracting more information. However, it is foreseeable that the accuracy of each dataset Libras, Appendicitis, Yale and WarpAR10P might increase when the parameter exceeds 1, which shows that the model FCADML is sensitive to δ.

Fig. 6 describes the trend of classification accuracy varying with the parameter β. It can be observed that accuracies of most selected datasets outperform well and remain stable in [0.05,0.35] except that the accuracy of dataset Phoneme firstly presents stable and then drops significantly. Besides, we also noted that three datasets Air, Wdbc and Appendicitis show a decreasing trend in accuracy varying from 0.35 to 0.45. Finally, from a macro perspective, it can be seen that the accuracy performance is relatively stable with a minor impact from β.

Next, we continue to study the influence of parameters δ and β in FCADML to verify the classification performance. The detailed accuracy results are presented in Fig. 7, from which we can confirm that the accuracies mostly increase with the increase of parameters δ and β.

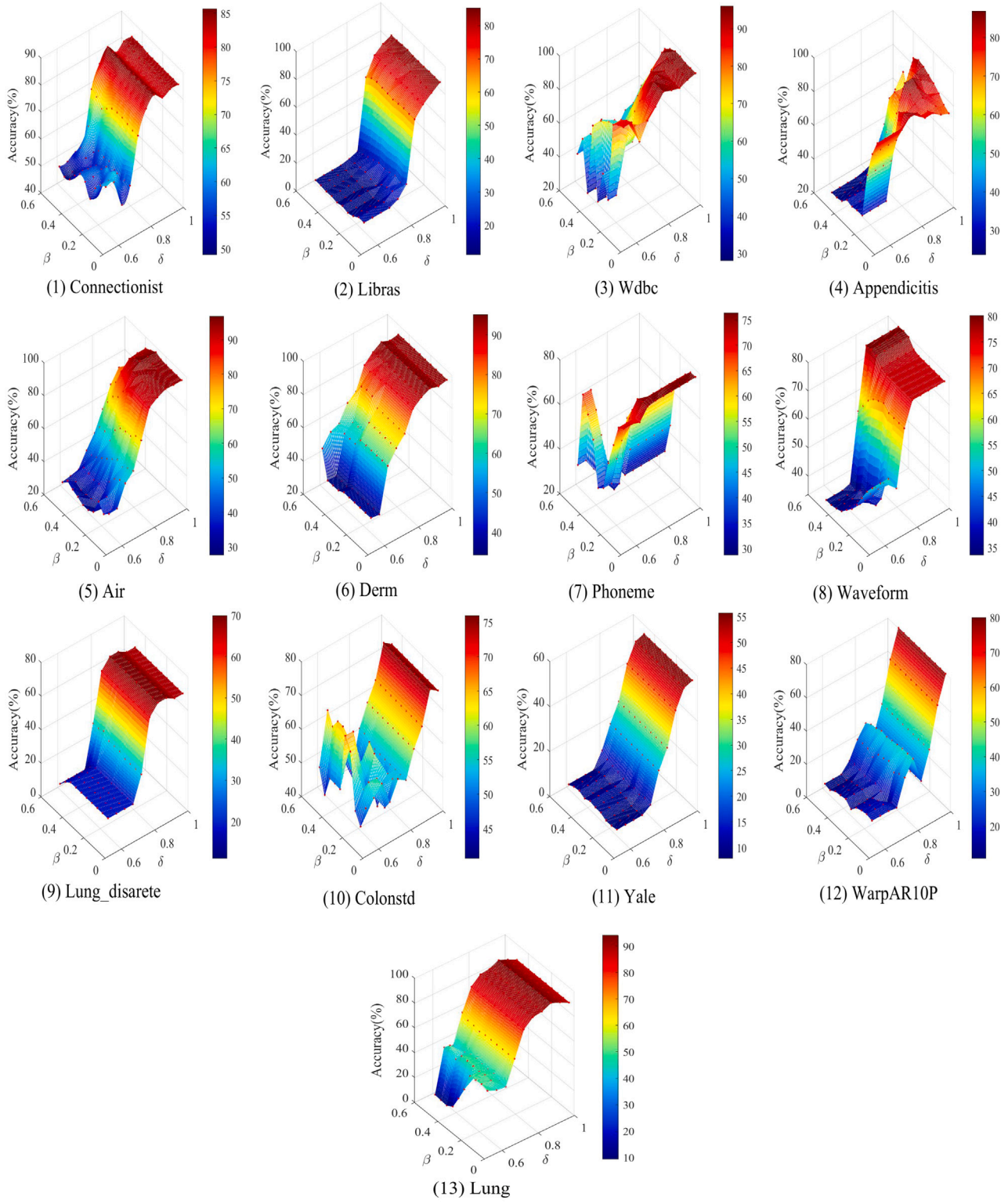

Fig. 7. Accuracy comparison as parameters δ and β on thirteen datasets.

It also can be indicated that the parameter β is not sensitive to classification accuracy and maintain accuracy fluctuations within a small range when δ is a fixed value in [0.8,1.0] excepting for datasets Wdbc, Appendicitis and Phoneme. Consequently, algorithm FCADML displays more sensitive to δ. It is essential to select an optimal parameter to improve the classification performance.

4.4. Analyses of statistical results

Additionally, to further evaluate the statistical significance among all fifteen algorithms, Friedman test [49] and Bonferroni–Dunn test [50] are identified whether there is an apparent difference with respect to the performance of selected models across thirteen datasets. A Fisher

Table 9
Rank of classification algorithms.

ID	FCADML	DMPWFC	ILMPFTC	FCLM	PIFWKNN	BM-FKNN	S3OFIS	FSMBGD	CT	CART	RSS	DT	Boosting	KNN	GSVM
1	4	2	1	5	3	15	6	9	12	10	14	13	11	7	8
2	1	12	2	8	7	11	5	6	9	10	13	14	15	4	3
3	2.5	14	9	15	5	2.5	7	6	10	11	13	12	8	4	1
4	1	12	10	13	3	14	15	9	5	8	6	11	7	4	2
5	1	14	3	12	6	7	8	4	10	9	15	13	11	2	5
6	2	10	6	15	8	11	7	5	3	9	14	13	12	1	4
7	8	9	13	11	3	14	15	10	4	2	12	6	7	1	5
8	5	14	8	11	3	13	15	2	7	9	6	12	10	4	1
9	7	1.5	1.5	3	6	15	5	8	11	9	10	14	12	4	13
10	4	10.5	10.5	14	9	15	5.5	8	3	1	13	5.5	2	7	12
11	6	3.5	3.5	7	5	14	2	1	10	9	11	13	12	8	15
12	1	7.5	7.5	10	9	15	6	5	3	2	11.5	13	11.5	4	14
13	5.5	2.5	2.5	7	1	15	4	10	8	11	14	9	12	5.5	13
Ave.	**3.69**	8.65	5.96	10.08	5.23	12.42	7.73	6.38	7.31	7.69	11.73	11.42	10.04	4.27	7.38

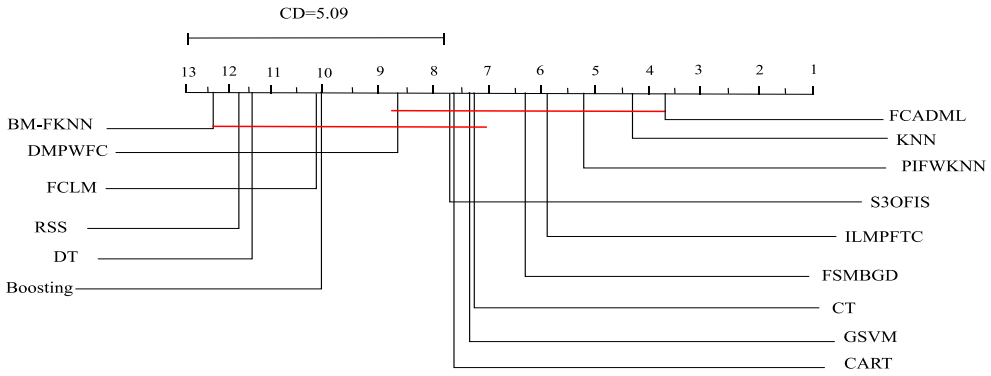

Fig. 8. CD comparison of all classification algorithms with Bonferroni–Dunn test ($\alpha = 0.05$).

distribution F_F of Friedman test is termed as:

$$F_F = \frac{(N-1)\chi_F^2}{N(k-1)-\chi_F^2} \sim F\big(k-1,(k-1)(N-1)\big), \tag{18}$$

in which $\chi_F^2 = \frac{12N}{k(k+1)}\big(\sum_{i=1}^{k} R_i^2 - \frac{k(k+1)^2}{4}\big)$. Then N and k represent the cardinalities of datasets and different algorithms, respectively. $R_i = \frac{1}{N}\sum_{j=1}^{N} r_j^i$ is the average rank of the ith algorithm across all datasets, here, r_j^i is the rank of the ith algorithm on the jth dataset. Initially, it is supposed that there are no obvious differences among all algorithms, and in fact, if $F_F > F\big(k-1,k-1(N-1)\big)$, then the original hypothesis is refuted. Actually, the rank results are described in Table 9 based on the cardinalities of algorithms. We can get $\chi_F^2 = 66.1138$ from the average rank of Table 9, and then χ_F^2 is put into Eq. (17) to obtain $F_F = 6.8461 > F(14,168) = 1.75$ at level $\alpha = 0.05$. Moreover, the null hypothesis does not hold and we adopt the alternative hypothesis that there exists a remarkable difference in the performance of fifteen classification algorithms.

In addition, what we are currently focusing on whether there is difference between any two classification algorithms, so we use Bonferroni–Dunn test to evaluate their statistical results. Then the critical difference is delineated as:

$$CD_\alpha = q_\alpha\sqrt{\frac{k(k+1)}{6N}} \tag{19}$$

where q_α represents the critical value at the significance level. Subsequently, the critical difference is $CD_\alpha = 5.09$ where $k = 15$ at level $\alpha = 0.05$. From Fig. 8, it is evident that FCADML has a significantly superior performance than BM-FKNN, FCLM, RSS, DT and Boosting. Meanwhile, the Bonferroni–Dunn test is not sufficient to demonstrate that there exist any significant differences among FCADML, KNN, PIFWKNN, S3OFIS, ILMPFTC, FSMBGD, CT, GSVM and CART. From CCL perspective, there is no obvious difference about performance between FCADML and ILMPFTC, while FCADML is better than DMPWFC and FCLM.

5. Conclusion

This article have explored an innovative association-based concept-cognitive learning method with distance metric learning for knowledge fusion and concept classification. The representative attribute set from attribute clusters is discussed to decrease the less relevant attributes and interfering information. Based on this, we propose a fuzzy concept associative learning model with distance metric learning, which is designed to learn fuzzy concepts that are closest to the clues, thus simulating the cognitive process of the human brain. Furthermore, concept clustering from the standpoint of pseudo-concept can compress the valid fuzzy concept associative space. At last, the performance superiority of the proposed model FCADML is verified from the accuracy and statistical results when compared with the existing CCL, fuzzy classification and machine learning models.

The introduction of associative learning in this article has opened up a new thought for concept-cognitive learning, which demonstrates the breadth and infinity of cognition. This idea breaks the limitations of constructing the concept space through the existing granular concept or neighborhood concept and masters the rich information for classification tasks. However, it is complex to construct the fuzzy concept associative space from formula (12) in terms of both time consumption and space storage when the clue is much abundant. Although we have provided a simple method for associative learning in formula (13), it is uncertain whether granular concept and neighborhood concept are the most valuable ones since they are only a small part of formula (12). Therefore, we will continue to focus on concept-cognitive associative learning to enhance its interpretability and application in the future research.

CRediT authorship contribution statement

Chengling Zhang: Writing – review & editing, Writing – original draft, Validation, Software, Conceptualization. **Guangming Xue:** Investigation, Data curation. **Weihua Xu:** Validation, Supervision, Funding

acquisition, Formal analysis. **Huilai Zhi:** Visualization, Software. **Yinfeng Zhou:** Visualization, Software. **Eric C.C. Tsang:** Visualization, Supervision, Funding acquisition.

Declaration of competing interest

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

Acknowledgments

This work is supported by the National Natural Science Foundation of China (No. 62376229), and Natural Science Foundation of Chongqing, China (No. CSTB2023NSCQ-LZX0027), and Natural Science Foundation of Fujian Province, China (2024J01793), Macao Science and Technology Development Funds (0036/2023/RIA1, and 0037/2024/RIA1), and the Natural Science Basic Research Program of Shaanxi Province, China (Grant No. 2024JC-YBQN-0025).

Data availability

Data will be made available on request.

References

[1] Z. Pylyshyn, Computation and cognition: Issues in the foundations of cognitive science, Behav. Brain Sci. 3 (1) (1980) 111–132.
[2] Y. Wang, N. Howard, J. Kacprzyk, et al., Cognitive informatics: Towards cognitive machine learning and autonomous knowledge manipulation, Int. J. Cogn. Inform. Nat. Intell. 12 (1) (2018) 1–13.
[3] J. Feldman, Minimization of boolean complexity in human concept learning, Nature 407 (6804) (2000) 630–633.
[4] B. Ganter, R. Wille, Formal Concept Analysis: Mathematical Foundations, Springer, New York, 1999.
[5] R. Wille, Restructuring lattice theory: An approach based on hierarchies of concepts, in: Formal Concept Analysis: 7th International Conference, ICFCA 2009 Darmstadt, Germany, May (2009) 21-24 Proceedings 7, Springer Berlin Heidelberg, 2009, pp. 314–339.
[6] M. Cintra, H. Camargo, M. Monard, Genetic generation of fuzzy systems with rule extraction using formal concept analysis, Inform. Sci. 349 (2016) 199–215.
[7] M. Fan, S. Luo, J. Li, Network rule extraction under the network formal context based on three-way decision, Appl. Intell. 53 (5) (2023) 5126–5145.
[8] C. Zhang, E. Tsang, W. Xu, et al., Dynamic updating variable precision three-way concept method based on two-way concept-cognitive learning in fuzzy formal contexts, Information Sciences 655 (2024) 119818.
[9] K. Yuan, W. Xu, W. Li, et al., An incremental learning mechanism for object classification based on progressive fuzzy three-way concept, Inform. Sci. 584 (2022) 127–147.
[10] C. Zhang, E.C.C. Tsang, W. Xu, et al., Incremental concept-cognitive learning approach for concept classification oriented to weighted fuzzy concepts, Knowl.-Based Syst. 260 (2023) 110093.
[11] Z. Pawlak, Rough sets, Internat. J. Comput. Inform. Sci. 11 (1982) 341–356.
[12] M. Prasad, S. Tripathi, K. Dahal, An efficient feature selection based bayesian and rough set approach for intrusion detection, Appl. Soft Comput. 87 (2020) 105980.
[13] K. Yuan, D. Miao, Y. Yao, et al., Feature selection using zentropy-based uncertainty measure, IEEE Trans. Fuzzy Syst. 32 (4) (2024) 2246–2260.
[14] M. Shao, Y. Leung, X. Wang, et al., Granular reducts of formal fuzzy contexts, Knowl.-Based Syst. 114 (2016) 156–166.
[15] M. Shao, H. Yang, W. Wu, Knowledge reduction in formal fuzzy contexts, Knowl.-Based Syst. 73 (2015) 265–275.
[16] W. Zhang, W. Xu, Cognitive model based on granular computing, Chin. J. Eng. Math. 24 (6) (2007) 957–971.
[17] Y. Yao, Interpreting concept learning in cognitive informatics and granular computing, IEEE Trans. Syst. Man Cybern. B 39 (4) (2009) 855–866.
[18] S. Yahia, K. Arour, A. Slimani, A. Jaoua, Discovery of compact rules in relational databases, Inform. Sci. 4 (3) (2000) 497–511.
[19] H. Zhi, J. Qi, T. Qian, et al., Three-way dual concept analysis, Internat. J. Approx. Reason. 114 (2019) 151–165.
[20] W. Xu, W. Li, Granular computing approach to two-way learning based on formal concept analysis in fuzzy datasets, IEEE Trans. Cybern. 46 (2) (2014) 366–379.
[21] W. Xu, D. Guo, Y. Qian, et al., Two-way concept-cognitive learning method: a fuzzy-based progressive learning, IEEE Trans. Fuzzy Syst. 31 (6) (2022) 1885–1899.
[22] J. Li, C. Mei, W. Xu, et al., Concept learning via granular computing: A cognitive viewpoint, Inform. Sci. 298 (2015) 447–467.
[23] Y. Shi, Y. Mi, J. Li, et al., Concept-cognitive learning model for incremental concept learning, IEEE Trans. Syst. Man Cybern.: Syst. 51 (2) (2018) 809–821.
[24] Y. Mi, Y. Shi, J. Li, et al., Fuzzy-based concept learning method: Exploiting data with fuzzy conceptual clustering, IEEE Trans. Cybern. 52 (1) (2020) 582–593.
[25] Y. Shi, Y. Mi, J. Li, et al., Concurrent concept-cognitive learning model for classification, Inform. Sci. 496 (2019) 65–81.
[26] W. Xu, Y. Chen, Multi-attention concept-cognitive learning model: A perspective from conceptual clustering, Knowl.-Based Syst. 252 (2022) 109472.
[27] X. Deng, J. Li, Y. Qian, et al., An emerging incremental fuzzy concept-cognitive learning model based on granular computing and conceptual knowledge clustering, IEEE Trans. Emerg. Top. Comput. Intell. 8 (3) (2024) 2417–2432.
[28] Y. Ding, W. Xu, W. Ding, Y. Qian, IFCRL: Interval-intent fuzzy concept re-cognition learning model, IEEE Trans. Fuzzy Syst. 32 (6) (2024) 3581–3593.
[29] D. Guo, W. Xu, Fuzzy-based concept-cognitive learning: An investigation of novel approach to tumor diagnosis analysis, Inform. Sci. 639 (2023) 118998.
[30] D. Guo, W. Xu, Y. Qian, et al., M-FCCL: Memory-based concept-cognitive learning for dynamic fuzzy data classification and knowledge fusion, Inf. Fusion 100 (2023) 101962.
[31] E. Xing, M. Jordan, S. Russell, et al., Distance metric learning with application to clustering with side-information, in: Proceedings of Advances in Neural Information Processing Systems, 2003, pp. 521–528.
[32] Z. Guo, Y. Ying, Guaranteed classification via regularized similarity learning, Neural Comput. 26 (3) (2014) 497–522.
[33] J. Yang, H. Xu, Metric learning based object recognition and retrieval, Neurocomputing 190 (2016) 70–81.
[34] H. Xiong, X. Chen, Kernel-based distance metric learning for microarray data classification, Bmc Bioinform. 7 (2006) 1–11.
[35] Y. Mu, W. Ding, D. Tao, Local discriminative distance metrics ensemble learning, Pattern Recognit. 46 (8) (2013) 2337–2349.
[36] K. Weinberger, L. Saul, Distance metric learning for large margin nearest neighbor classification, J. Mach. Learn. Res. 10 (2) (2009) 207–244.
[37] A. Michel, J. Farrell, Associative memories via artificial neural networks, IEEE Control Syst. Mag. 10 (3) (1990) 6–17.
[38] D. Bassett, M. Mattar, A network neuroscience of human learning: Potential to inform quantitative theories of brain and behavior, Trends Cogn. Sci. 21 (4) (2017) 250–264.
[39] N. Biswas, S. Chakraborty, S. Mullick, et al., A parameter independent fuzzy weighted k-nearest neighbor classifier, Pattern Recognit. Lett. 101 (2018) 80–87.
[40] M. Kumbure, P. Luukka, M. Collan, A new fuzzy k-nearest neighbor classifier based on the Bonferroni mean, Pattern Recognit. Lett. 140 (2020) 172–178.
[41] X. Gu, An explainable semi-supervised self-organizing fuzzy inference system for streaming data classification, Inform. Sci. 583 (2022) 364–385.
[42] Y. Cui, D. Wu, J. Huang, Optimize tsk fuzzy systems for classification problems: Minibatch gradient descent with uniform regularization and batch normalization, IEEE Trans. Fuzzy Syst. 28 (12) (2020) 3065–3075.
[43] I. Witten, E. Frank, M. Hall, et al., Data Mining: Practical Machine Learning Tools and Techniques, vol. 2, no. 4, Morgan Kaufmann, 2005.
[44] T. Cover, P. Hart, Nearest neighbor pattern classification, IEEE Trans. Inform. Theory 13 (1) (1967) 21–27.
[45] G. McLachlan, Discriminant Analysis and Statistical Pattern Recognition, John Wiley & Sons, 2005.
[46] S. Keerthi, C. Lin, Asymptotic behaviors of support vector machines with Gaussian kernel, Neural Comput. 15 (7) (2003) 1667–1689.
[47] D. Cutler, J. Edwards, K. Beard, et al., Random forests for classification in ecology, Ecology 88 (11) (2007) 2783–2792.
[48] L. Breiman, Bagging predictors, Mach. Learn. 24 (1996) 123–140.
[49] M. Friedman, A comparison of alternative tests of significance for the problem of M rankings, Ann. Math. Stat. 11 (1) (1940) 86–92.
[50] Q. Dunn, Multiple comparisons among means, J. Amer. Statist. Assoc. 56 (293) (1961) 52–64.