AIMS Mathematics

http://www.aimspress.com/journal/Math

AIMS Mathematics, 10(9): 22127–22149.
DOI: 10.3934/math.2025985
Received: 05 July 2025
Revised: 05 September 2025
Accepted: 08 September 2025
Published: 24 September 2025

Research article

Fuzzy concept cognitive learning based on knowledge space theory

Ju Huang[1,2], **Yidong Lin**[1,2,*] **and Wen Sun**[1,2]

[1] School of Mathematics and Statistics, Minnan Normal University, Zhangzhou 363000, China
[2] Fujian Key Laboratory of Granular Computing and Applications, Minnan Normal University, Zhangzhou 363000, China

* **Correspondence:** Email: yidong_lin@yeah.net.

Abstract: Concept cognitive learning (CCL) constitutes a rigorous and current cognitive theory for the representation and learning of concepts of the human brain. The well-established CCL models pay close attention to construct a concept space, in which all the attributes are mastered concurrently. However, few attempts have been made to combine CCL with cognitive logic in a fuzzy context due to attribute precedence. For this case, this paper first develops a cognitive surmise relationship among attributes, cognitive transitions, and discrimination of fuzzy concepts based on knowledge space theory. Furthermore, utilizing the inherent information of concepts, the attributes are weighted to accurately understand and apply fuzzy concepts. To better derive benefits from the fuzziness and uncertainty of knowledge, an approach is provided to improve performance through the fusion of fuzzy concepts. Empirical studies on twenty datasets reveal the effectiveness and efficiency of the proposed model.

Keywords: concept cognitive learning; concept fusion; fuzzy concept; knowledge space theory
Mathematics Subject Classification: 03B52, 03E72

1. Introduction

Concept-cognitive learning (CCL) simulates human learning mechanisms to incorporate new data into themselves to adapt the dynamic environments. It was first studied in [42] from an abstract perspective. A concept is a key pillar of CCL systems and the base unit of cognition. It is identified by its extent and intent along with a Galois from formal concept analysis (FCA) [28]. Accordingly, CCL is a new learning theory based on FCA as its theoretical foundation. The learned societies strenuously developed the fuzzy concept [37], three-way concept [24], multi-scale concept [4], approximate concept [13], and AFS concept [30] to satisfy different situations. Generally speaking, the framework of CCL [18] includes concept constructing and mapping it into different sub-concept spaces. Then, it takes the concept space as a knowledge storage carrier. In addition, the concept space is collections of the granular concept or incomplete concept from a context. In general, the study of CCL includes three different aspects: 1.1 Two-way concept learning; 1.2 Continual learning; and 1.3 Multi-level CCL.

1.1. Two-way concept learning

As we known, CCL was first studied from an abstract perspective in [42]. Then Xu et al. [12, 31, 32] launched detail analysis for carefully addressing the two-way translation of information granule concepts, and a fuzzy-based CCL [33] system in a dynamic situation was further studied. From the concept movement viewpoint, the three-way decision (3WD) was integrated with two-way learning to delve deeper into the concept evolution [35]. Meanwhile, emphasizing the latent skills capable of solving problems from knowledge space theory [8], Xie et al. [36] developed a novel two-way learning technique semantically for attribute-oriented concepts.

1.2. Continual learning

Concept continual learning aims at continuously learning and adapting to new concepts in a constantly changing environment. Li et al. [14] first integrated granular computing into a cognitive concept lattice and established a cognitive computing system by integrating granular computing into a cognitive concept lattice and axiomatic approaches for three-way CCL via multi-granularity [15]. Then, Zhao et al. [43] explored an approximate cognitive computing system for incomplete information. To address the large-scale datasets, a map reduction framework for granular CCL [19] and a concurrent incremental learning technique [25] were respectively developed. In addition, to tackle complex learning tasks and reduce uncertainty, a fuzzy-based CCL method [20], incremental cognition of concepts [6, 16, 17, 26], progressive fuzzy three-way CCL method [41], multi-attention CCL method [34], and CCL for concept drift and decision making [22] have been well-established.

1.3. Multi-level CCL

Multi-level CCL emphasizes the establishment of deep concept networks within a cognitive process. Yao [38] proposed a conceptual framework based on a layered model of knowledge discovery. Fan and Tsang [11, 27] constructed attribute-oriented and feature-oriented multi-level CCLs to recognize certain objects and distinguish them rather than identifying all objects. Zhang et al. proposed attribute topology [44] and the incremental concept tree [45], which clearly displayed the relationship between new data and original data. Yan et al. [39, 40] brought forward the construction of three-way attribute partial order structure and the incremental CCL algorithm successively. Mi [23] pioneered the proposal of concept neural networks grounded in the concept space.

To summarize, there are several shortcomings in the emerging mainstream CCLs. In particular, 1) *Lack of learning smoothness.* When learning new knowledge or skills, it is best to organize the learning content in increasing levels of difficulty and complexity. 2) *Lack of reasonable weighting mechanism.* Different concepts play different roles in the learning process. 3) *Acquire a major expenditure of time and effort for CCL.* In practice, the search space is filled with a large number of similar fuzzy concepts or similar labeled fuzzy concepts, which can be fused as legitimately as possible. These are the main challenges in our work and will be innovatively overcome. In this paper, a novel fuzzy-based CCL is proposed for fuzzy formal (decision) contexts, including surmise relationships and transfers of attributes, mechanisms of fuzzy concepts with weights and fusion.

The remainder of this paper is organized as follows. Section 2 briefly reviews some basic notions about fuzzy formal contexts and knowledge spaces. Section 3 presents the theory of cognitive logic, such as the surmise relation and cognitive inference from a fuzzy formal context. Section 4 establishes a novel cognition mechanism based on cognitive logic and concept fusion. The experimental results on some real datasets are reported in Section 5. Finally, this paper is concluded with several challenges for further research in Section 6.

2. Preliminaries

In this section, we briefly review some basic notions about fuzzy concepts and knowledge spaces.

The triple $(U, A, \tilde{I})$ is called a fuzzy formal context, in which U is the domain of the objects, A is the collection of attributes, and $\tilde{I}$ is the fuzzy relationship between U and A. For $X \subseteq U$ and $\tilde{B} \in \mathcal{F}(A)$, $\mathcal{F}(A)$ is the fuzzy power set of A, and the concept operators $f : \mathcal{P}(U) \to \mathcal{F}(A)$ and $g : \mathcal{F}(A) \to \mathcal{P}(U)$ are respectively defined as follows:

$$f(X)(a) = \bigwedge_{x \in X} \tilde{I}(x, a), a \in A, \tag{2.1}$$

$$g(\tilde{B}) = \{x \in U \mid \forall a \in A, \tilde{B}(a) \le \tilde{I}(x, a)\}. \tag{2.2}$$

Then a pair $(X, \tilde{B})$ is called a fuzzy concept if $f(X) = \tilde{B}$ and $g(\tilde{B}) = X$. The collection of all fuzzy concepts $L(U, A, \tilde{I})$ forms a complete lattice that is called the fuzzy concept lattice of $(U, A, \tilde{I})$.

The quintuple $(U, A, \tilde{I}, D, J)$ is called a fuzzy formal decision context where $(U, A, \tilde{I})$ is a fuzzy formal context, $J : U \times D \to \{0, 1\}$, $A \cap D = \emptyset$, and $D = \{d_1, d_2, \cdots, d_l\}$ is a set of decision attributes. If for any $d_1, d_2 \in D$, $\mathcal{H}(\{d_1\}) \cap \mathcal{H}(\{d_2\}) = \varnothing$ holds, $(U, A, \tilde{I}, D, J)$ is called a regular fuzzy formal decision context. The partition of U by D is denoted as $U/D = \{U_i \mid i = 1, 2, \cdots, l\}$. Furthermore, (U, D, J) is termed a decision formal context. In CCL, according to decision classes, the regular fuzzy formal decision context is usually split into several sub-contexts to generate the concept space for clue cognition.

By Doignon and Falmagne [8], a knowledge structure is a pair $(Q, \mathcal{K})$ (or $\mathcal{K}$ for short) in which Q is a domain and $\mathcal{K}$ is a collection of subsets of Q, containing at least the empty set $\emptyset$ and domain Q. When a knowledge structure $(Q, \mathcal{K})$ is closed under union, we call $(Q, \mathcal{K})$ a knowledge space. Furthermore, a knowledge space closed under intersection is called a quasi ordinal space. There is a one-to-one correspondence between the collection of all quasi ordinal spaces $\mathcal{K}$ on a domain Q, and the collection of all quasi orders $\mathcal{R}$ on Q [1]. Such a correspondence is defined by, for $x, y \in Q$,

$$x\mathcal{R}y \Leftrightarrow (\forall K \in \mathcal{K},\ y \in K \Rightarrow x \in K). \tag{2.3}$$

Accordingly, $L(U, A, \tilde{I})$ is a quasi ordinal space.

Let $R \subseteq U \times U$. If $R^2 \subseteq R$, then R is referred to as a transitive relation, where for any $(x, y) \in U \times U$,

$$R^2(x, y) = (R \circ R)(x, y) = \bigvee_{z \in U} (R(x, z) \wedge R(z, y)). \tag{2.4}$$

The transitive closure of R is the minimum transitive relation containing R, that is,

$$t(R) = \bigcup_{k=1}^{\infty} R^k. \tag{2.5}$$

3. Cognitive preference

An important part of this section focuses on analyzing possible ways to cognize the material within cognitive concepts. This leads naturally to studying the "predecessor" of an attribute. Intuitively, an attribute a is a predecessor of an attribute b if a is never mastered after b, either for logical or historical reasons. However, we are looking for the logical or cognitive reasons for this among all the attributes. To be more specific, the following formalizes the intuitive ideas in fuzzy scene.

Assumption 1. *As the difficulty of knowledge increases, the number of learners gradually decreases.*

Assumption 2. *The cognitive level of a successor is not much higher than that of the predecessor in the case of careless or lucky guesses.*

Assumption 3. *The current cognitive state is only related to the cognitive state of the previous moment.*

Note the example from education where easy problems are easier to master than difficult ones. In this case, the number of learners decreases with the increase in the difficulty of problems. In the case of careless or lucky guesses, the scores for two questions with different levels of difficulty are kept within a certain range. Cognitive learning also satisfies the Markov property.

Definition 1. *Let $(U, A, \tilde{I})$ be a fuzzy formal context, and a predecessor function $h : A \times A \to 2^U$ is defined by, for $(a, b) \in A \times A$,*

$$h(a, b) = \{x \in U \mid \tilde{I}(x, a) \geq \tilde{I}(x, b)\}. \tag{3.1}$$

Obviously, $h(a, a) = U$ and $h(a, b) \neq h(b, a)$ in normal circumstances.

Theorem 1. *Let $(U, A, \tilde{I})$ be a fuzzy formal context with a predecessor function $h : A \times A \to 2^U$. For $a, b, c \in A$,*

1. *$h(a, a) = U$,*
2. *$h(a, b) \cap h(b, c) \subseteq h(a, c)$,*
3. *$h^c(a, b) \cap h^c(b, c) \subseteq h^c(a, c)$, where $h^c(a, b)$ is the complement of $h(a, b)$.*

Proof. By definition, for $x \in h(a, b) \cap h(b, c)$ we have $\tilde{I}(x, a) \geq \tilde{I}(x, b)$ and $\tilde{I}(x, b) \geq \tilde{I}(x, c)$. Then $x \in h(a, c)$, namely $h(a, b) \cap h(b, c) \subseteq h(a, c)$. For $y \in h^c(a, b) \cap h^c(b, c)$, we have $\tilde{I}(y, a) < \tilde{I}(y, b)$ and $\tilde{I}(y, b) < \tilde{I}(y, c)$, and then $y \in h^c(a, c)$. That is, $h^c(a, b) \cap h^c(b, c) \subseteq h^c(a, c)$. □

With Assumption 1, fault tolerance is naturally considered when lucky guesses and careless errors happen. Thus, in the following, we first propose a surmise degree of cognitive attributes to realize this idea.

Definition 2. *Let $(U, A, \tilde{I})$ be a fuzzy formal context. The surmise degree $S : A \times A \to [0, 1]$ is defined by, for $a, b \in A$,*

$$S(a, b) = \frac{|h(a, b)|}{|U|}. \tag{3.2}$$

Theorem 2. *Let $(U, A, \tilde{I})$ be a fuzzy formal context. For $a, b, c \in A$, the following properties hold:*

1. *$S(a, a) = 1$,*

2. $S(a,b) = S(b,a)$ *if and only if* $|h(a,b)| = |h(b,a)|$,
3. *if* $S(a,b) \geq \delta$, $S(b,c) \geq \delta$, $\delta \in (0,1]$, *then* $S(a,c) \geq \delta$.

Proof. Only item (3) is proved.

$$S(a,c) = \frac{|h(a,c)|}{|U|} \geq \frac{|h(a,b) \cap h(b,c)|}{|U|} = 1 - \frac{|h^c(a,b) \cup h^c(b,c)|}{|U|}$$
$$= 1 - \frac{|h(a,b)|}{|U|} - \frac{|h(b,c)|}{|U|} + 2\frac{|h(a,b) \cap h(b,c)|}{|U|} - \frac{|[h(a,b) \cup h(b,c)]^c|}{|U|}.$$

Since $S(a,b) \geq \delta$ and $S(b,c) \geq \delta$, then

$$\frac{|[h(a,b) \cup h(b,c)]^c|}{|U|} \leq 1 - \delta.$$

In this case,

$$S(a,c) \geq S(a,b) + S(b,c) + \frac{|[h(a,b) \cup h(b,c)]^c|}{|U|} - 1 \geq \delta.$$

□

The collection of surmise degrees for any pair of attributes would form a surmise degree matrix, denoted by M_S. Therefore, the element of the ith row crosses the jth column $M_S(i,j) = S(a_i, a_j)$ for $a_i, a_j \in A$. With Assumption 2, we next consider a cognitive error tolerance as follows.

Definition 3. *Let* $(U, A, \tilde{I})$ *be a fuzzy formal context, and a cognitive error tolerance* $T : A \times A \to [0,1]$ *is defined for* $a, b \in A$ *by*

$$T(a,b) = \begin{cases} max\left\{\tilde{I}(x,b) - \tilde{I}(x,a) \mid x \in h^c(a,b)\right\}, & a \neq b \\ 0 & , \quad a = b \end{cases}. \tag{3.3}$$

Generally, $T(a,b) \neq T(b,a)$ if $a \neq b$ for $a, b \in A$. Such a cognitive error tolerance can also be presented as a matrix M_T, in which $M_T(a,b) = T(a,b)$.

Example 1. *Table 1 is a fuzzy formal context* $(U, A, \tilde{I})$*, in which* $U = \{x_1, x_2, x_3, x_4\}$, $A = \{a_1, a_2, a_3, a_4, a_5\}$. *From Table 1, by Definitions 2 and 3, the surmise degree matrix is:*

$$M_S = \begin{bmatrix} 1 & 0.5 & 1 & 0.75 & 0.75 \\ 0.75 & 1 & 1 & 0.5 & 0.75 \\ 0.75 & 0.25 & 1 & 0.5 & 0.5 \\ 0.5 & 0.5 & 0.5 & 1 & 1 \\ 0.25 & 0.5 & 0.5 & 0 & 1 \end{bmatrix}.$$

The cognitive error tolerance is

$$M_T = \begin{bmatrix} 0 & 0.1 & 0 & 0.7 & 0.1 \\ 0.2 & 0 & 0 & 0.6 & 0.1 \\ 0.7 & 0.5 & 0 & 0.7 & 0.6 \\ 0.5 & 0.6 & 0.5 & 0 & 0 \\ 0.6 & 0.6 & 0.6 & 0.6 & 0 \end{bmatrix},$$

in which $M_T(a_i, a_j) = T(a_i, a_j)$, $i, j = 1, 2, 3, 4, 5$.

Table 1. A formal fuzzy context $(U, A, \tilde{I})$.

$\tilde{I}$	a_1	a_2	a_3	a_4	a_5
x_1	0.9	0.7	0.2	0.9	0.8
x_2	0.8	0.8	0.8	0.3	0.2
x_3	0.1	0.2	0.1	0.8	0.2
x_4	0.7	0.8	0.7	0.2	0.2

Theorem 3. *For fuzzy formal context $(U, A, \tilde{I})$, $T : A \times A \to [0, 1]$ is the corresponded cognitive error tolerance. For $a, b, c \in A$, $T(a, c) \leq T(a, b) \vee T(b, c)$.*

Proof. As we know, $T(a, b) = \bigvee_{x \in h^c(a,b)} \left(\tilde{I}(x, b) - \tilde{I}(x, a)\right)$ with $a \neq b$ for $a, b \in A$. On the other hand, $h^c(a, c) \subseteq h^c(a, b) \cup h^c(b, c)$ by Theorem 1. Consequently, $T(a, c) \leq \left[\bigvee_{x \in h^c(a,b)} (\tilde{I}(x, b) - \tilde{I}(x, a))\right] \vee \left[\bigvee_{x \in h^c(b,c)} (\tilde{I}(x, c) - \tilde{I}(x, b))\right] = T(a, b) \vee T(b, c)$. □

Corollary 1. *For $a, b, c \in A$ and parameter β, if $T(a, b) \leq \beta$ and $T(b, c) \leq \beta$, where T is a cognitive error tolerance with respect to fuzzy formal context $(U, A, \tilde{I})$, then $T(a, c) \leq \beta$.*

Definition 4. *Let $(U, A, \tilde{I})$ be a fuzzy formal context, and let $R \subseteq A \times A$ be a relation on A defined for $a, b \in A$ by*

$$(a, b) \in R \Leftrightarrow S(a, b) \geq \delta \text{ and } T(a, b) < \beta, \tag{3.4}$$

where $\delta \in (0.5, 1]$ and $\beta \in [0, 0.5]$ are control parameters. R is called a (δ,β)-surmise relation on A.

Theorem 4. *Let $(U, A, \tilde{I})$ be a fuzzy formal context and R be a (δ,β)-surmise relation on A. For $a, b, c \in A$,*

1. *$(a, a) \in R$,*
2. *if $(a, b) \in R$, $(b, c) \in R$, then $(a, c) \in R$.*

Proof. Since $S(a, a) = 1$ and $T(a, a) = 0$ by Definitions 2 and 3 for $a \in A$, then $(a, a) \in R$. If $(a, b) \in R$ and $(b, c) \in R$, then $S(a, b) \geq \delta$, $S(b, c) \geq \delta$, $T(a, b) \leq \beta$, and $T(b, c) \leq \beta$. According to item (3) of Theorem 2 and Theorem 3, $S(a, c) \geq \delta$ and $T(a, c) \leq \beta$ obviously hold. Fusing these two conclusions, we have $(a, c) \in R$. □

Thus, R is reflexive and transitive, and a quasi order relation naturally. Note that we denote

$$M_S^\delta(i, j) = \begin{cases} 1, & M_S(i, j) \geq \delta \\ 0, & otherwise \end{cases},$$

$$M_T^\beta(i, j) = \begin{cases} 1, & M_T(i, j) < \beta \\ 0, & otherwise \end{cases}.$$

Comparatively, M_S^δ and M_T^β generate a (δ,β)-surmise relation matrix $M_R^{(\delta,\beta)} = M_S^\delta \wedge M_T^\beta$. For convenience, we call M_S^δ, M_T^β, and $M_R^{(\delta,\beta)}$ the δ-surmise matrix, β-tolerance matrix and (δ,β)-reachability matrix respectively.

Definition 5. *Let R_1 and R_2 be two (δ,β)-surmise relations on A loaded from fuzzy formal concept $(U, A, \tilde{I})$. We say that $R_1 \leq R_2$ iff for all $(a, b) \in R_1$ implies $(a, b) \in R_2$ for $a, b \in A$.*

Theorem 5. *Let $(U, A, \tilde{I})$ be a fuzzy formal context. Let R_1 and R_2 be a (δ_1, β_1)-surmise relation and (δ_2, β_2)-surmise relation, respectively. If $\delta_1 \leq \delta_2$ and $\beta_2 \leq \beta_1$, then $R_2 \leq R_1$.*

Proof. If $(a, b) \in R_2$, then $S(a, b) \geq \delta_2 \geq \delta_1$ and $T(a, b) < \beta_2 \leq \beta_1$. This means $(a, b) \in R_1$. Hence, $R_2 \leq R_1$. □

Definition 6. *For a (δ, β)-surmise relation R on A in $(U, A, \tilde{I})$, the (δ, β)-adjacency relation R^* on A is defined by, for $a, b, c \in A$, $(a, b) \in R^*$ if and only if $(a, b) \in R \wedge (a \neq b) \wedge [(a, c) \in R \wedge (c, b) \in R \Rightarrow b = c]$.*

We denote $M_{R^*}^{(\delta,\beta)}$ as the corresponding (δ, β)-adjacency matrix of R^*, which is usually sparse.

Example 2. *Take $\delta = 0.75$ and $\beta = 0.5$. The $(0.75, 0.5)$-surmise relation matrix and $(0.75, 0.5)$-adjacency matrix are*

$$M_R^{(0.75,0.5)} = \begin{bmatrix} 1 & 0 & 1 & 0 & 1 \\ 1 & 1 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}, \ M_{R^*}^{(0.75,0.5)} = \begin{bmatrix} 0 & 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

Theorem 6. *Let R and R^* be the (δ, β)-surmise relation and (δ, β)-adjacency relation on A with respect to $(U, A, \tilde{I})$, and let E be an $|A| \times |A|$ identity matrix. Then*

$$M_R^{(\delta,\beta)} = t(M_{R^*}^{(\delta,\beta)} + E). \tag{3.5}$$

Proof. Suppose $M_R^{(\delta,\beta)} = (r_{ij})_{|A|\times|A|}$ and $M_{R^*}^{(\delta,\beta)} + E = (r^*_{ij})_{|A|\times|A|}$. For $i, j = 1, 2, \cdots, |A|$, we clearly have

- $r_{ij}^{*(2)} = \bigvee_{j_1=1}^{|A|} (r^*_{ij_1} \wedge r^*_{j_1 j})$,
- $r_{ij}^{*(3)} = \bigvee_{j_1=1}^{|A|} \bigvee_{j_2=1}^{|A|} (r^*_{ij_1} \wedge r^*_{j_1 j_2} \wedge r^*_{j_2 j})$,
- $\vdots$
- $r_{ij}^{*(k)} = \bigvee_{j_1=1}^{|A|} \bigvee_{j_2=1}^{|A|} \cdots \bigvee_{j_{k-1}=1}^{|A|} (r^*_{ij_1} \wedge r^*_{j_1 j_2} \wedge \cdots \wedge r^*_{j_{k-1} j})$.

Apparently, a repetition will rise from such indexes $j_1, j_2, \cdots, j_{k-1}, j$ with respect to $r_{ij}^{*(k)}$ whenever $k > |A|$. In such a case, there exists $p \leq |A|$ such that

$$r_{ij}^{*(k)} = \bigvee_{j_1=1}^{|A|} \bigvee_{j_2=1}^{|A|} \cdots \bigvee_{j_{k-1}=1}^{|A|} (r^*_{ij_1} \wedge r^*_{j_1 j_2} \wedge \cdots \wedge r^*_{j_{k-1} j}) = r_{ij}^{*(p)} \leq \bigvee_{p=1}^{|A|} r_{ij}^{*(p)}.$$

Namely, $\left(M_{R^*}^{(\delta,\beta)} + E\right)^k \subseteq \bigcup_{p=1}^{|A|} \left(M_{R^*}^{(\delta,\beta)} + E\right)^p$ for any $k > |A|$. Then

$$t\left(M_{R^*}^{(\delta,\beta)} + E\right) = \bigcup_{k=1}^{\infty} \left(M_{R^*}^{(\delta,\beta)} + E\right)^k = \left[\bigcup_{k=1}^{p} (M_{R^*}^{(\delta,\beta)} + E)^k\right] \cup \left[\bigcup_{k=p+1}^{\infty} (M_{R^*}^{(\delta,\beta)} + E)^k\right] = \bigcup_{k=1}^{p} \left(M_{R^*}^{(\delta,\beta)} + E\right)^k.$$

Therefore, $t(M_{R^*}^{(\delta,\beta)} + E) = \bigcup_{k=1}^{|A|} (M_{R^*}^{(\delta,\beta)} + E)^k$. Next, we elaborate thoroughly that $t\left(M_{R^*}^{(\delta,\beta)} + E\right) = \left(M_{R^*}^{(\delta,\beta)} + E\right)^k$ for $k \geq |A|$. Obviously, $M_{R^*}^{(\delta,\beta)} + E$ is reflexive and so $r^*_{ii} = 1$ for $1 \leq i \leq |A|$. This indicates

$$r_{ij}^{*(2)} = \bigvee_{p=1}^{|A|} (r^*_{ip} \wedge r^*_{pj}) \geq r^*_{ii} \wedge r^*_{ij} = r^*_{ij},$$

that is, $M_{R^*}^{(\delta,\beta)} + E \subseteq \left(M_{R^*}^{(\delta,\beta)} + E\right)^2$. More normally, we have

$$\left(M_{R^*}^{(\delta,\beta)} + E\right)^{p+1} = \left(M_{R^*}^{(\delta,\beta)} + E\right)^p \circ \left(M_{R^*}^{(\delta,\beta)} + E\right) \subseteq \left(M_{R^*}^{(\delta,\beta)} + E\right)^p \circ \left(M_{R^*}^{(\delta,\beta)} + E\right)^2 = \left(M_{R^*}^{(\delta,\beta)} + E\right)^{p+2}$$

for $p \geq 1$. That is to say, $M_{R^*}^{(\delta,\beta)} + E \subseteq \left(M_{R^*}^{(\delta,\beta)} + E\right)^2 \subseteq \cdots \subseteq \left(M_{R^*}^{(\delta,\beta)} + E\right)^{|A|} \subseteq \cdots \subseteq \left(M_{R^*}^{(\delta,\beta)} + E\right)^k \subseteq \cdots$. Thus, $t\left(M_{R^*}^{(\delta,\beta)} + E\right) = \bigcup_{p=1}^{|A|} \left(M_{R^*}^{(\delta,\beta)} + E\right)^p \subseteq \left(M_{R^*}^{(\delta,\beta)} + E\right)^k$. On the other hand, we have $\left(M_{R^*}^{(\delta,\beta)} + E\right)^k \subseteq \bigcup_{p=1}^{|A|} \left(M_{R^*}^{(\delta,\beta)} + E\right)^p = t\left(M_{R^*}^{(\delta,\beta)} + E\right)$. Then $t\left(M_{R^*}^{(\delta,\beta)} + E\right) = \left(M_{R^*}^{(\delta,\beta)} + E\right)^k$ for $k \geq |A|$. Last, we only have to proof $M_R^{(\delta,\beta)} = (M_{R^*}^{(\delta,\beta)} + E)^{|A|}$. The reachability matrix $M_R^{(\delta,\beta)}$ having significant reflexivity and transitivity implies $t\left(M_R^{(\delta,\beta)}\right) = M_R^{(\delta,\beta)}$. We can check that $M_{R^*}^{(\delta,\beta)} + E \subseteq M_R^{(\delta,\beta)}$, i.e., $r_{ij}^* \leq r_{ij}$ for $i, j \in \{1, 2, \cdots, |A|\}$, which indicates $t\left(M_{R^*}^{(\delta,\beta)} + E\right) \subseteq M_R^{(\delta,\beta)}$. In what follows, $r_{ij}^{*(p)} \geq r_{ij}$ will be authenticated against the (δ,β)-adjacency relation. If $r_{ij} = 0$, then $r_{ij}^{*(p)} \geq r_{ij}$ evidently, $a_i, a_j \in A$. If $r_{ij} = 1$, $(a_i, a_j) \in R$. Next, two cases need to be analyzed. The first case is $(a_i, a_j) \in R^*$. Then $r_{ij}^{*(p)} = r_{ij}$. The another case is $(a_i, a_j) \notin R^*$. Then there exists a sequence $\{a_{j_1}, a_{j_2}, \cdots, a_{j_p}\} \subseteq A$ such that $(a_i, a_{j_1}) \in R^*$, $(a_{j_1}, a_{j_2}) \in R^*, \cdots$, $(a_{j_p}, a_j) \in R^*$. Accordingly, $r_{ij_1}^* = 1$, $r_{j_{k-1}j_k}^* = 1$, $2 \leq k \leq p$, and $r_{j_p j}^* = 1$. These indicate $r_{ij}^{*(p)} = 1$, $p \leq |A|$. Then it follows that $r_{ij}^{*(p)} = 1$ no matter if $p \leq |A|$ or $p \geq |A|$. Therefore $t\left(M_{R^*}^{(\delta,\beta)} + E\right) = M_R^{(\delta,\beta)}$ holds. □

With Assumption 3, the attribute cognition in current time only relates legitimately to the previous one, a Markov process.

Definition 7. *Let* $(U, A, \tilde{I})$ *be a fuzzy formal context with a* (δ,β)*-surmise relation* R*,* $\delta \in (0.5, 1]$ *and* $\beta \in [0, 0.5]$*. The inner fringe and outer fringe of a cognitive attribute* $a \in A$ *are respectively defined as*

$$a^{\mathcal{I}} = \{b \in A \mid (b, a) \in R^*\}, \quad a^{\mathcal{O}} = \{b \in A \mid (a, b) \in R^*\}.$$

Definition 8. *When* $(a, b) \in R^*$ *for* $a, b \in A$*, the cognitive transfer measure is defined by*

$$Ct(a \to b) = -\frac{1}{|\mathcal{G}|} \sum_{(X,\tilde{B})\in\mathcal{G}} \tilde{B}(b) log \tilde{B}(a). \tag{3.6}$$

Beyond question, this cross entropy is useful information to characterize the transition probability between a pair of attributes.

Example 3. *According to Example 2, the cognitive transitions depicted in Figure 1, we see that* $Ct(a_2 \to a_1) = 0.186$*,* $Ct(a_1 \to a_3) = 0.139$*,* $Ct(a_1 \to a_5) = 0.145$*, and* $Ct(a_4 \to a_5) = 0.202$.

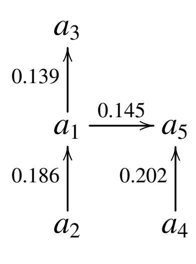


Figure 1. Hasse diagram of $(0.75, 0.5)$-surmise relation R with cognitive transitions.

Definition 9. *Let R be a (δ,β)-surmise relation related to a fuzzy formal context $(U, A, \tilde{I})$. For any $b \in A$, if $b^O = \emptyset$ and there is a finite sequence of attributes $a = p_0, p_1, \cdots, p_k = b$ such that $a^I = \emptyset$, $p_i \in p_{i-1}^O$, $1 \leq i \leq k$, we call the sequence of sets a tight cognitive path from a to b and denote it as $P(a \to b)$.*

Corollary 2. *If $(a, b) \in R$ with $a^I = b^O = \emptyset$ for $a, b \in A$, then there is a tight cognitive path from a to b.*

Predictably, there may be multiple cognitive paths from a to b if $a^I = b^O = \emptyset$ and $(a, b) \in R$. This depends in part on the significance of cognitive attributes. Thus, the significance of attributes is clarified by the following definition.

Definition 10. *Let $(U, A, \tilde{I})$ be a fuzzy formal context and R be the related (δ,β)-surmise relationship. For $a \in A$, define*

$$Sig(a) = \sum_{b \in a^O} Ct(a \to b) + \sum_{b \in a^I} Ct(b \to a) \tag{3.7}$$

as the significance of a.

Then from Figure 1, $Sig(a_1) = 0.47$, $Sig(a_2) = 0.186$, $Sig(a_3) = 0.139$, $Sig(a_4) = 0.202$, and $Sig(a_5) = 0.347$.

4. Concept learning for classification

In concept cognition and recognition, the global discrimination of concepts is first explored to fuse concepts by means of concept weighting to reduce the size of the concept space. Then there is only needed to learn the related knowledge for clues according to the new fuzzy concept space. We use Euclidean distance to depict the discrimination among distinct concepts, that is, for $(X, \tilde{B}), (Y, \tilde{C}) \in \mathcal{G}$,

$$dis((X, \tilde{B}), (Y, \tilde{C})) = \left[\sum_{a \in A} (\tilde{B}(a) - \tilde{C}(a))^2\right]^{1/2}. \tag{4.1}$$

Definition 11. *Let $\mathcal{G}$ be a concept space. For $(X, \tilde{B}) \in \mathcal{G}$,*

$$Dis((X, \tilde{B})) = \frac{\sum_{(Y,\tilde{C}) \in \mathcal{G}'} dis(\tilde{B}, \tilde{C})}{|\mathcal{G}'|} \tag{4.2}$$

is referred to as the discrimination of concept $(X, \tilde{B})$, in which $\mathcal{G}' = \{(Y, \tilde{C}) \in \mathcal{G} | dis(\tilde{B}, \tilde{C}) \leq \theta\}$ and θ is a parameter. We denote it as $Dis(\tilde{B})$ for short without confusion.

Obviously, the higher $dis(\tilde{B}, \tilde{C})$ is, the greater the difference between them and less correlation there is between them. $Dis(\tilde{B})$ depicts the global discrimination of $(X, \tilde{B})$, which actually reflects the degree of dispersion of concepts in the neighborhood centered on $(X, \tilde{B})$.

Definition 12. *Let $\mathcal{G}$ be a concept space. For a clue $(x, \tilde{B})$, we call*

$$LA\left(x, (Y, \tilde{C})\right) = dis\left(\tilde{C} * Sig, \tilde{B} * Sig\right) - r\left(\mathcal{G}'\right) \tag{4.3}$$

the learning accuracy of x with respect to $(Y, \tilde{C}) \in \mathcal{G}$, where

$$Sig = \left(Sig(a_1), Sig(a_2), \cdots, Sig(a_{|A|})\right),$$

$$\mathcal{G}' = \left\{(Z, \tilde{K}) \in \mathcal{G} \mid dis(\tilde{C}, \tilde{K}) \leq r(\mathcal{G}')\right\},$$

$$r(\mathcal{G}') = \frac{1}{|\mathcal{G}'|} \sum_{(P,\tilde{Q}) \in \mathcal{G}'} dis\left(\tilde{Q} * Sig, \frac{1}{|\mathcal{G}'|} \sum_{(Z,\tilde{K}) \in \mathcal{G}'} \tilde{K} * Sig\right),$$

in which "$*$" means the Hadamard product of matrices.

Thus, learning accuracy is positively correlated with discrimination and negatively correlated with the radius of the neighborhood. Then the learning objective for a given clue is minimizing

$$CL(x, (Y, \tilde{C})) = \frac{LA\left(x, (Y, \tilde{C})\right)}{Dis\left(\frac{1}{|\mathcal{G}'|} \sum_{(Z,\tilde{K}) \in \mathcal{G}'} \tilde{K} * Sig\right)}, \tag{4.4}$$

and the parameter θ in Definition 12 is now the sphere radius of $\mathcal{G}'$.

Let $(U, A, \tilde{I}, D, J)$ be a regular fuzzy formal decision context, where $U/J = \{D_t, D_2, \cdots, D_t\}$ is a partition of D, and $J \subseteq U \times D$. For $x \in D_i$ $(1 \leq i \leq t)$, its neighborhood [41] is constructed in the following:

$$N_x = \{y \in D_i \mid d(x, y) \leq \alpha\}, \tag{4.5}$$

in which $d(\cdot)$ denotes the Euclidean distance and α is a control parameter. [41] gives the following definition of degraded version of the fuzzy concept space.

Definition 13. *$(U, A, \tilde{I}, D, J)$ is a regular fuzzy formal decision context, where $U/J = \{D_t, D_2, \cdots, D_t\}$. For $D_i \in U/J$, the fuzzy concept space $\mathcal{S}_i$ related to D_i is*

$$\mathcal{S}_i = \{(g \circ f(N_x), f(N_x)) \mid x \in D_i\}. \tag{4.6}$$

Then the collection $\mathcal{S}$ of all S_i, $1 \leq i \leq t$, is called a fuzzy concept space, where $\mathcal{S}_i$ is said to be a subspace of $\mathcal{S}$.

Let $Sig = (Sig(a_1), Sig(a_2), \cdots, Sig(a_{|A|}))$ be the weight of attributes. Then the fusion mechanism with respect to $\mathcal{S}_i$ is

$$\tilde{C} = \frac{\sum_{(Z,\tilde{K}) \in \mathcal{S}_i} \tilde{K} * Sig}{|\mathcal{S}_i|}, \tag{4.7}$$

$$r(\mathcal{S}_i) = \frac{\sum_{(Z,\tilde{K}) \in \mathcal{S}_i} dis(\tilde{K} * Sig, \tilde{C})}{|\mathcal{S}_i|}. \tag{4.8}$$

In this case, $\tilde{C}$ is the description (equivalent in intention) of the fusion center, and $r(\mathcal{S}_i)$ is the radius. Afterward, the global discrimination of subspace $\mathcal{S}_i$ is built up on the fusion as below.

$$Dis(\mathcal{S}_i) = \frac{\sum_{(Z,\tilde{K}) \in \mathcal{S}_i} dis(\tilde{C}, \tilde{K})}{|\mathcal{S}_i|}. \tag{4.9}$$

Subsequently, for a given clue $(x, \tilde{B})$, the learning accuracy of x with respect to $\mathcal{S}_i$ is regularized by the following equation:

$$LA(x, \mathcal{S}_i) = dis\left(\tilde{C} * Sig, \tilde{B} * Sig\right) - r(\mathcal{S}_i), \tag{4.10}$$

where "$*$" is the Hadamard product of matrices. Then an algorithm CCDS for concept classification on regular fuzzy formal decision contexts with the given clue is available as follows.

Algorithm 1 Concept classification based on discrimination and significance (CCDS)

Input: A regular fuzzy formal decision context $(U, A, \tilde{I}, D, J)$ and clue $(x, \tilde{B})$.
Output: The decision label of x.
1: Compute the significance of $a \in A$ and fuzzy concept space $\mathcal{S}$.
2: Compute the fusion center, radius, and discrimination of $\mathcal{S}_i \in \mathcal{S}$.
3: Compute $LA(x, \mathcal{S}_i)$.
4: Find $\mathcal{S}' = arcmax_{\mathcal{S}_i \in \mathcal{S}} LA(x, \mathcal{S}_i)$.
5: If $|\mathcal{S}'| = 1$, output the corresponding decision label.
6: Else if $|\mathcal{S}'| > 1$, then output the decision label related to the subspace satisfying $arcmin_{\mathcal{S}_i \in \mathcal{S}'} Dis(\mathcal{S}_i)$.

The actual expense of Algorithm 1 is far less than $O(\sum_{i=1}^{|U/J|} |D_i| + |A|^2 + |A|)$ and usually close to $O(\sum_{i=1}^{|U/J|} |D_i| + |A|)$ since the adjacency matrix is sparse. However, S2CL [21] needs $O(E \times \sum_{i=1}^{|U/J|}(|D_i| + |D_i|^2))$, in which E means the step size of incremental learning. ILMPFTC [41] has a time consumption of $O(|A| \times \sum_{i=1}^{|U/J|} \sum_{j=1}^{|D_i|}(3 + 2(|D_i| - |g \circ f(N_{x_j}) \cap g^c \circ f^c(N_{x_j})|)))$ and DMPWFC [46] takes $O(\sum_{i=1}^{|U/J|}(|U/J|(|D_i|^2 + |w\mathcal{S}_i|) + \sum_{j=1}^{|D_i|} |A|(1 + (|U| - x_j^{**}))))$, where g^c and f^c denote the negative cognitive operators [41], $w\mathcal{S}_i$ is the weighted concept space, and x_j^{**} is a collection of objects that dissatisfy with $w\mathcal{S}_i$ [46]. We can prove that CCDS has the smallest time complexity.

5. Experiments

In this section, we empirically evaluate the effectiveness of CCDS on concept classification. Specifically, CCDS is compared with several mainstream CCL methods and traditional machine learning classification algorithms through implementation on various datasets from the UCI* repository. The experiments are independently implemented 10 times with random data partitions in MATLAB 2020a and carried out on a personal computer with an Intel Core(TM) i7-9700 @3.00GHz CPU and 8.00GB main memory, and organized through the following three aspects: (1) parametric analysis of CCDS; (2) the classification performance and efficiency of CCDS; and (3) the statistical significance test of CCDS.

5.1. Experimental design

To verify the performance of the proposed model (CCDS), three well-established models in CCL are selected for comparison: S2CL [21], ILMPFTC [41], and DMPWFC [46]. Additionally. several classic models from machine learning are also included in the comparative experiments: K-nearest neighbors (KNN) [3], decision tree (DT) [29], naive bayes (NB) [10], classification and regression tree (CART) [2], SVM†, and neural networks (NN‡). There is a common parameter α as the radius of the neighborhood in CCDS, ILMPFTC, and DMPWFC. Therefore, α is optimally selected to classify

*http://archive.ics.uci.edu/datasets
†https://ww2.mathworks.cn/help/stats/fitcsvm.html
‡https://ww2.mathworks.cn/help/stats/fitcnet.html

accuracy and ranged in $[0, 1]$ with step 0.05. The parameter of KNN is set as $k = 3$. In CCDS, δ is chosen from $0.5, 0.55, ..., 1$ and β on $0, 0.05, ..., 0.5$ with step 0.05. In the experiments, 20 datasets are selected and presented in Table 2. Each dataset is randomly split into two parts—-training data and testing data–which account for 70% and 30% of the data, respectively. In the preprocessing stage, the dataset is fuzzified to get the membership degree belonging to the interval [0,1]. Therefore, these datasets are first fuzzified by utilizing [7],

$$\tilde{R}(x_i, a_j) = \frac{f(x_i, a_j) - min(f(a_j))}{max(f(a_j)) - min(f(a_j))}, \tag{5.1}$$

where $f(x_i, a_j)$ denotes the value of x_i with respect to attribute a_j, and $max(f(a_j))$ and $min(f(a_j))$ are the maximum and minimum of all objects with respect to a_j, respectively. The fuzzy value of $\tilde{R}(x, a)$ means the membership degree of (x, a) to $\tilde{R}$ in the fuzzy formal decision context. The fuzzy set $\tilde{R}$ can be understood as the degree of membership of objects to attributes. Thus the greater $\tilde{R}(x, a)$ is, the greater the degree to which x owns a, and Eq 5.1 as a method of fuzzification could convert the experimental data into a fuzzy formal decision context.

Table 2. Detailed information on experimental data.

ID	Dataset	Training	Testing	Feature	Class
1	Soy	33	14	21	4
2	Iris	105	45	4	3
3	Thyroid	151	64	5	3
4	Heart	189	81	13	2
5	Ionosphere	246	105	34	2
6	Derm	257	109	34	6
7	Wdbc	398	171	30	2
8	Balance	438	187	4	3
9	Crx	483	207	15	2
10	Austrlian Credit	483	207	14	2
11	Wiscon	489	210	9	2
12	Pima	538	230	8	2
13	German	700	300	20	2
14	Sick	1960	840	29	2
15	Ablone	2924	1253	8	3
16	Spam	3221	1380	57	2
17	Wilt	3387	1452	5	2
18	Waveform	3500	1500	21	3
19	Mushroom	5687	2437	22	2
20	Gamma	13314	5706	10	2

5.2. Parametric analysis of CCDS

We analyze three parameters α, β, and δ on 20 datasets depicted in Table 2, and show the experimental results in Figures 2–5. Figure 2 shows the average classification accuracy influenced by α, from which we observe that the experiments achieve the optimal effect for most datasets when α is employed in [0.0,0.3]. In addition, the accuracy changes dramatically in some cases. When $\alpha = 0.525$, the upward fluctuations of the results of several datasets are more obvious. Thus, our method is extremely sensitive to α. Figures 3 and 4 respectively show the accuracy trends of parameters β and δ. It can be seen that CCDS performs well on most datasets in Figure 3 when β is selected in [0,0.3]. Furthermore, for most datasets, we can also identify that the accuracy gradually deteriorates when β changes from 0.35 to 0.5 in several cases. Figure 4 reflects that the performance remains relatively stable on most of the datasets affected by δ. In other words, our method has better generalization performance with respect to δ.

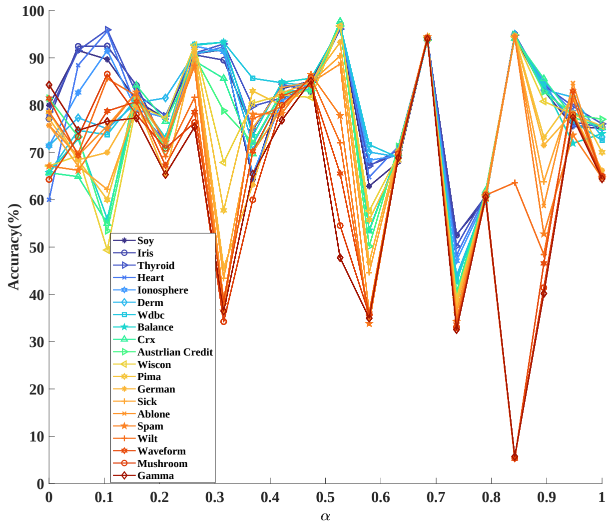


Figure 2. The influence of α.

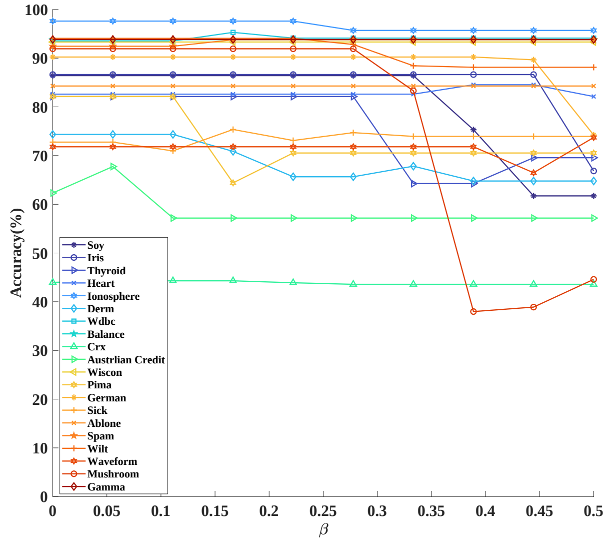


Figure 3. The influence of β.

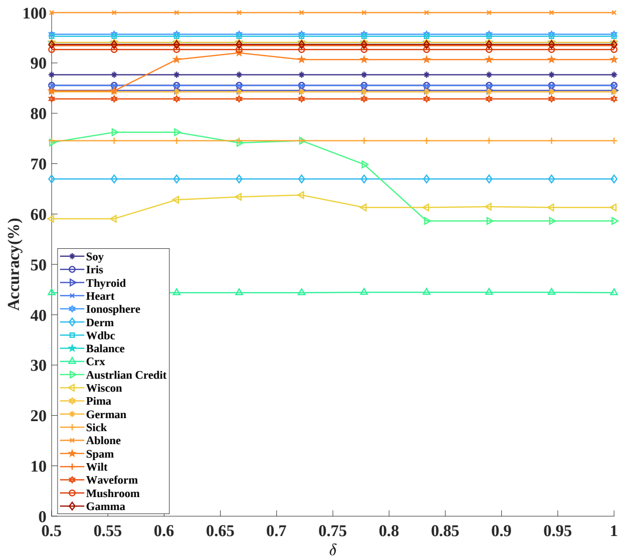
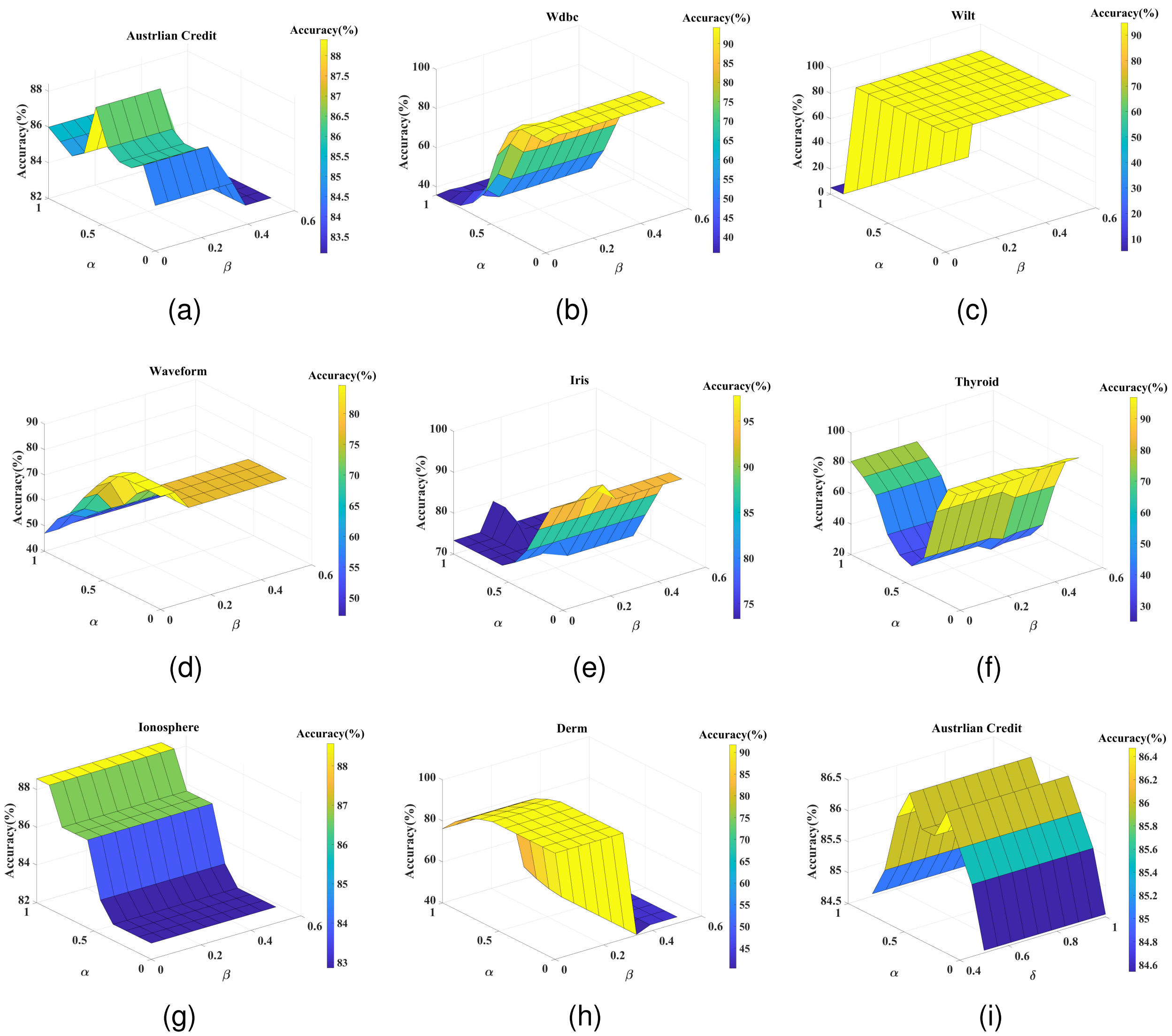
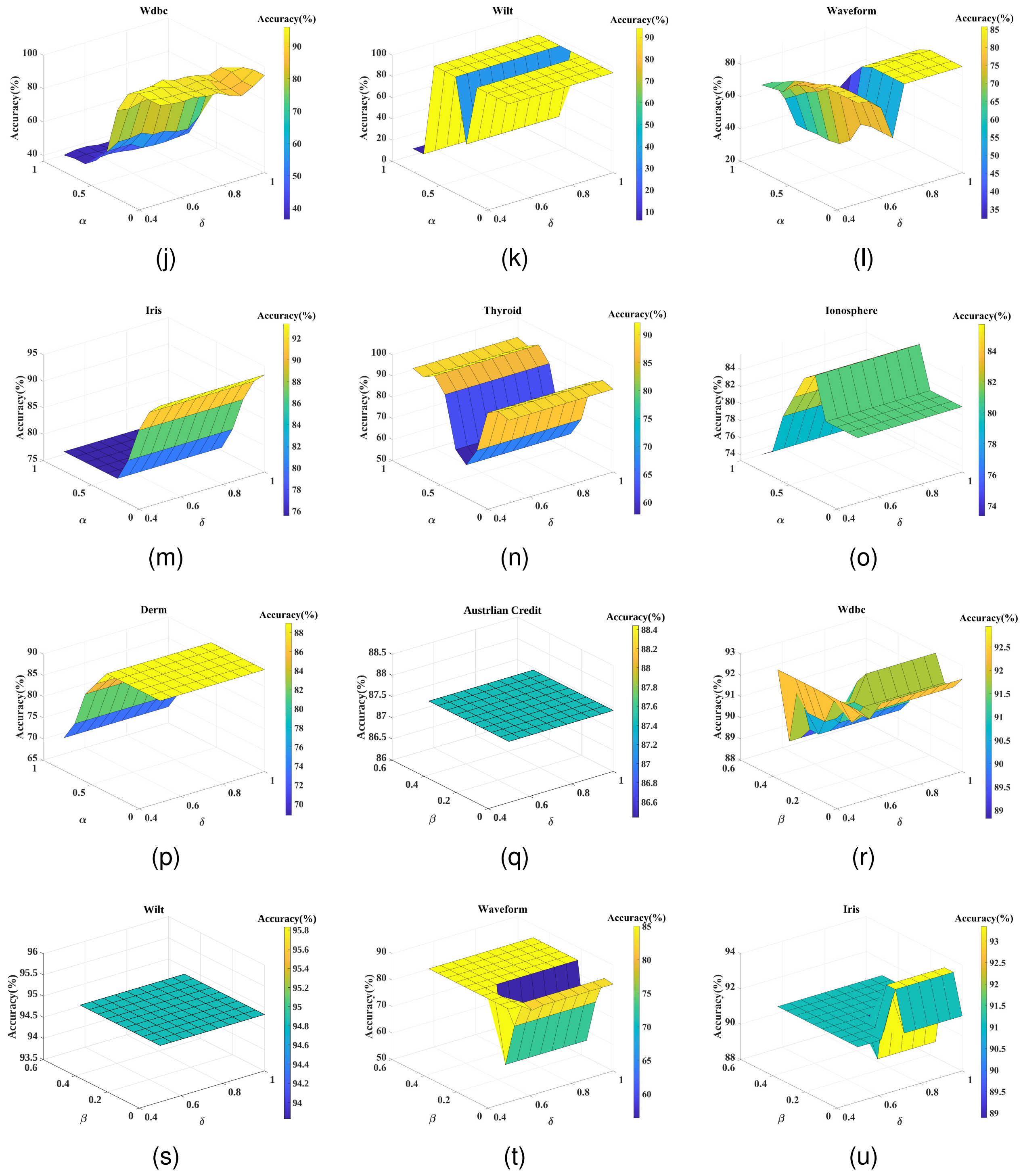


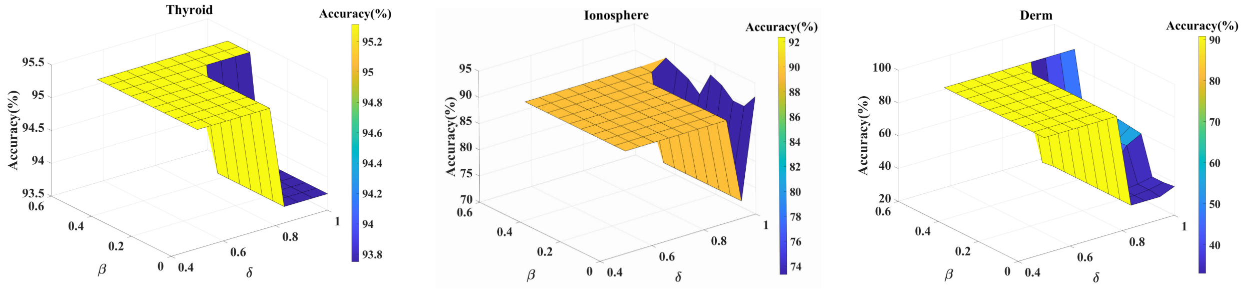
Figure 4. The influence of δ.

To further evaluate the performance influenced by each pair of parameters in CCDS, as representatives, we analyze these parameter pairs on the Austrlian Credit, Wdbc, Wilt, Waveform, Iris, Thyroid, Ionosphere and Derm datasets. There is a similar phenomenon with the remaining datasets. Figure 5 shows the average classification result by comparing the pairwise combinations of parameters α, β, and δ. Figure 5(a)–(h) visualize the trend of accuracy with (α, β), while Figure 5(i)–(p) and (q)–(x) visualize that of accuracy with (α, δ), and (β, δ), respectively. It is clear that most datasets are insensitive to β and δ in Figure 5(a)–(x). However, the accuracy decreases when α increases. As an example, from Figure 5(a)–(g), (i)–(o), and (r), we can see that an insufficient influence of β as well as δ on the classification results exists. More specifically, when $\beta \leq 0.4$ is employed, the slope of the accuracy plane remains constant along the direction of β in Figure 5(b)–(h), (q), (s), and (u)–(x). In conclusion, the proposed model is sensitive to α, and unresponsive to β and δ.





Wdbc
Accuracy(%)
α
δ
(j)
Wilt
(k)
Waveform
(l)
Iris
(m)
Thyroid
(n)
Ionosphere
(o)
Derm
(p)
Austrlian Credit
β
(q)
Wdbc
(r)
Wilt
(s)
Waveform
(t)
Iris
(u)



(v) (w) (x)

Figure 5. The accuracy by comparing α, β, and δ pairwise.

6. Classification performance and efficiency of CCDS

To evaluate the performance and efficiency, we focus on four key metrics for each method: accuracy, standard deviation, F1-score, and running time. We record the average results across 20 datasets for comparison, as shown in Tables 3 to 7.

From Tables 3 to 5, we can draw a couple of observations. (1) Compared with the concept cognitive methods, CCDS achieves better performance on 15 out of 20 datasets with respect to accuracy, while on datasets 5 and 14, both of them rank second and their gaps with the corresponding champions are both very small. (2) On all 20 datasets, CCDS is superior to other cognitive algorithms with respect to the average accuracy. This suggests that exploiting the attribute surmise relationship is conducive to performance improvement. (3) These comparative methods achieve the best performance on up to 5 out of 20 datasets. (4) On running time, our method (CCDS) outperforms all others. Particularly on datasets 14 to 19, its time consumption does not exceed one second, which is about 30 times less than that of S2CL, the best comparative method. While processing on large scale data (e.g., Dataset 20), the consumption gap has been further widened. According to F1-score, CCDS wins on 11 out of 20 datasets, while on Dataset 5, it ranks second with a small gap between itself and the champion. Accordingly, CCDS is significantly superior to the others. This demonstrates that reducing search space sensibly contributes to the efficiency and performance improvement. Thus, we conclude that the proposed method is effective for CCL, and has advantages compared with some other well-established CCLs.

To further demonstrate the performance of our method, we conduct six traditional classifications including KNN, CART, DT, NB, SVM, and NN on all datasets, as shown in Tables 6 and 7. In Table 6, CCDS achieves better performance on 7 out of 20 datasets, while KNN, CART, DT, NB, SVM, and NN achieve better performance on 0, 4, 2, 3, 2, and 4 out of 20 datasets, respectively. Table 7 reports that CCDS performs better on 7 out of 20 datasets, while the compared methods perform better on only 1, 3, 3, 2, 3, and 4 out of 20 datasets respectively in terms of F1-score. Meanwhile, only CCDS did not have the classification errors on all datasets. This shows that our method is among the best.

Table 3. Comparison results of CCL methods (mean±standard deviation%) in terms of accuracy.

ID	CCDS	ILMPFTC	S2CL	DMPWFC
1	**100.00±0.00**	100±0.00	100.00±0.00	100.00±0.00
2	**97.78±4.56**	94.29±2.13	95.24±2.92	76.19±3.37
3	**98.43±1.77**	97.10±1.35	97.10±2.65	95.97±3.42
4	**88.89±3.61**	75.59±2.23	73.67±3.40	77.85±9.85
5	88.57±7.48	86.54±3.26	**89.23± 2.67**	87.98±3.40
6	**96.33±3.22**	95.28±1.33	96.23±1.33	94.81±3.34
7	**97.06±1.78**	94.59±1.34	96.00±0.77	91.47±0.42
8	**89.84±4.16**	65.59±1.78	71.72±3.93	75.05±0.20
9	**86.95±2.40**	81.36±1.70	79.90±3.59	77.91±0.34
10	**90.82±2.19**	80.49±3.35	80.78±2.13	82.77±1.03
11	**99.04±1.19**	95.89±0.80	96.27±0.40	95.22±0.68
12	**76.95±4.05**	68.91±1.70	70.66±2.65	66.59±5.25
13	**71.67±2.09**	68.76±1.92	69.23±2.62	68.06±0.24
14	95.24±0.79	**95.92±0.62**	94.92±0.47	87.60±2.70
15	**54.35±1.44**	51.21±0.72	35.30±4.32	27.16±14.68
16	76.88±10.93	87.90±0.50	**90.40±0.70**	39.38±0.00
17	**95.31±0.40**	94.90±0.34	79.27±1.36	51.07±25.34
18	**87.40±1.46**	76.92±0.74	77.25±1.03	77.22±1.18
19	90.85±6.95	**100.00±0.00**	99.91±0.05	51.77±0.00
20	75.88±0.36	51.79±1.07	**80.72±0.55**	80.39±0.20
Avg.Acc.	**87.91**	82.27	83.69	74.41

Table 4. Comparison results of CCL methods (mean deviation%) in terms of F1-score.

ID	CCDS	ILMPFTC	S2CL	DMPWFC
1	**100.00**	100	100.00	100.00
2	**98.08**	92.92	90.72	80.51
3	97.96	98.09	**100.00**	91.70
4	**88.68**	81.29	78.09	70.42
5	85.59	**86.90**	84.52	84.24
6	91.42	**96.45**	95.50	91.53
7	**96.95**	93.67	94.94	91.23
8	**60.00**	52.88	57.60	45.85
9	**86.60**	80.41	83.35	78.92
10	**90.41**	81.49	84.29	84.46
11	**98.96**	95.22	96.28	95.57
12	**73.84**	62.88	70.58	63.87
13	41.75	59.47	**65.34**	62.60
14	48.62	**76.88**	76.87	59.70
15	**55.54**	51.63	39.10	-
16	74.31	87.39	**90.06**	-
17	48.55	**69.08**	60.44	54.83
18	**87.66**	75.96	77.71	76.36
19	90.85	**100.00**	99.96	-
20	72.27	59.30	**79.50**	77.64

"-" means that this algorithm has all of the classification errors in a class.

Table 5. Running time(s) of the concept cognitive algorithms.

ID	$\|U\|$	CCDS	ILMPFTC	S2CL	DMPWFC
1	47	**0.01**	0.01	0.01	0.02
2	150	**0.01**	0.03	0.02	0.13
3	215	**0.01**	0.05	0.05	0.18
4	270	**0.01**	0.16	0.11	0.20
5	351	**0.02**	0.13	0.11	0.50
6	366	**0.03**	0.12	0.09	0.26
7	569	**0.03**	0.62	0.36	0.92
8	690	**0.02**	0.64	0.57	1.95
9	690	**0.03**	0.79	0.45	1.16
10	699	**0.03**	0.78	0.50	1.36
11	768	**0.04**	1.04	0.56	4.35
12	869	**0.03**	0.70	0.37	2.00
13	1000	**0.04**	0.71	0.68	2.22
14	2800	**0.30**	13.07	10.03	54.18
15	4177	**0.33**	27.87	9.07	246.01
16	4601	**0.65**	43.03	17.63	648.39
17	4839	**0.87**	73.13	38.53	408.57
18	5000	**0.59**	12.87	11.00	616.19
19	8124	**0.93**	50.22	42.10	459.51
20	19020	**4.24**	1546.59	977.90	1732.91

Table 6. Comparison with machine learning algorithms (mean±standard deviation%) in terms of accuracy.

ID	CCDS	KNN	CART	DT	NB	SVM	NN
1	**100.00±0.00**	95.50±3.19	98.00±2.11	98.00±2.11	88.00±10.52	95.50±3.19	98.00±2.11
2	**97.78±4.56**	90.00±1.89	93.33±1.81	94.67±2.30	10.67±1.78	94.67±1.41	94.67±1.75
3	**98.43±1.77**	95.35±2.09	97.19±1.53	96.75±1.03	95.35±1.79	97.71±1.08	96.32±1.20
4	**88.89±3.61**	78.52±2.96	74.81±3.23	74.81±2.31	82.96±1.67	80.74±3.48	78.15±2.43
5	88.57±7.48	86.31±1.79	87.74±1.57	**91.74±1.52**	83.21±1.74	84.63±1.73	89.17±1.95
6	96.33±3.32	94.82±1.08	**99.45±0.39**	95.08±1.13	97.52±1.11	98.63±0.65	98.08±0.61
7	**97.06±1.78**	90.41±1.47	96.69±0.59	93.28±0.83	89.28±1.43	91.53±1.03	96.79±1.18
8	**89.83±4.16**	76.67±1.68	81.01±1.50	81.88±1.78	86.23±1.07	85.80±0.85	84.78±1.23
9	**86.95±2.40**	77.97±1.20	79.28±2.09	82.46±1.79	85.65±1.13	84.93±1.27	83.77±0.78
10	90.82±2.19	90.27±1.17	91.99±1.29	91.41±1.15	**93.41±0.94**	92.85±0.92	89.83±0.98
11	99.04±1.19	98.44±0.34	**99.61±0.21**	98.70±0.46	95.44±0.87	99.09±0.36	98.96±0.34
12	76.95±4.05	98.24±0.48	**99.82±0.18**	98.94±0.50	99.30±0.30	**99.82±0.18**	99.65±0.25
13	71.67±2.09	72.00±1.18	67.90±1.84	66.90±0.87	**74.70±1.56**	73.70±0.96	69.60±1.18
14	95.24±0.79	95.39±0.46	95.36±0.35	**97.61±0.29**	96.36±0.31	93.89±0.34	96.07±0.43
15	54.35±1.44	48.86±0.53	48.46±1.30	48.48±0.72	**55.74±0.91**	52.53±0.58	55.50±1.03
16	76.88±10.93	93.83±0.51	93.65±0.20	94.72±0.37	94.04±0.39	94.87±0.28	**95.96±0.32**
17	95.31±0.40	95.27±0.34	97.48±0.23	96.71±0.18	94.71±0.24	94.61±0.35	**97.89±0.22**
18	87.40±1.46	97.06±0.22	96.92±0.28	97.28±0.25	96.10±0.30	**98.60±0.17**	98.50±0.13
19	90.85±6.95	99.99±0.01	**99.99±0.01**	99.99±0.01	99.79±0.16	98.33±0.16	**99.99±0.01**
20	75.88±0.36	99.89±0.02	99.99±0.01	99.93±0.02	99.77±0.04	99.99±0.01	**100.00±0.00**
Avg.Acc.	87.91	88.74	89.95	89.97	85.91	90.62	**91.08**

Table 7. Comparison with machine learning algorithms (mean deviation%) in terms of F1-score.

ID	CCDS	KNN	CART	DT	NB	SVM	NN
1	**100.00**	-	98.75	-	-	-	-
2	**98.08**	90.82	94.07	94.11	-	95.40	95.05
3	**97.96**	-	96.79	95.81	93.30	97.50	-
4	**88.68**	79.11	73.45	73.11	82.74	81.33	78.00
5	85.59	85.72	87.14	**91.17**	81.69	82.65	88.29
6	91.42	-	**99.38**	95.25	97.36	98.90	-
7	**96.95**	79.12	92.87	86.73	-	-	94.32
8	60.00	77.86	80.80	81.82	86.25	**86.36**	84.61
9	**86.60**	77.87	79.17	82.49	85.57	85.27	83.58
10	90.41	89.00	90.99	90.66	**92.79**	92.16	88.71
11	98.96	98.31	**99.56**	98.60	94.93	99.01	98.85
12	73.84	98.16	99.79	98.84	99.25	**99.81**	99.64
13	41.75	62.29	62.28	61.25	**66.44**	64.30	63.36
14	48.62	74.80	78.47	**89.19**	82.62	-	-
15	**55.54**	49.97	48.73	48.91	55.35	-	55.25
16	74.31	93.76	93.33	94.47	93.74	94.63	**95.75**
17	48.55	80.79	87.65	82.92	66.45	-	**89.49**
18	87.66	97.11	97.01	97.30	96.09	**98.60**	98.51
19	90.85	**99.99**	**99.99**	**99.99**	-	98.33	**99.99**
20	72.27	99.88	99.99	99.92	99.75	99.99	**100.00**

"-" means that this algorithm has all of the classification errors in a class.

Statistical significance test of CCDS

To further analyze the performance among all of the methods, the Friedman test [9] and Bonferroni-Dunn test [5] are used as the favorable statistical significance tests for the method comparison on the

20 datasets. For the Friedman test, a Fisher distribution F_F is denoted as follows:

$$F_F = \frac{(N-1)\mathcal{X}_F^2}{N(k-1) - \mathcal{X}_F^2}, \tag{6.1}$$

where $\mathcal{X}_F^2 = \frac{12N}{k(k+1)}\left(\sum_{i=1}^{k} R_i^2 - \frac{k(k+1)^2}{4}\right)$, N is the cardinal number of the datasets, k is the number of experimental models, and R_i means the average rank of the algorithms on all datasets. F_F is a Friedman statistic with $k-1$ and $(k-1)(N-1)$ degrees of freedom. If $F_F > F(k-1, (k-1)(N-1))$, then the null hypothesis is rejected. In fact, the null hypothesis, which follows the principle that all the methods have equal performance, is clearly rejected in terms of $F_F = 8.4494$ and the critical value of $F(9, 171) = 1.668$ at significance level $\alpha = 0.1$. Thus, the classification performance of ten models is remarkably different.

Furthermore, the post-hoc Bonferroni-Dunn test is utilized to complete the performance analysis. Here, CCDS is regarded as the control method whose average rank difference from the compared methods is calibrated with the critical difference (CD) via the following equation:

$$CD_\alpha = q_\alpha \sqrt{\frac{k(k+1)}{6N}}, \tag{6.2}$$

where q_α is denoted as the critical value in the test. Accordingly, CCDS is deemed to have significantly different performance to one compared model if their average ranks differ by at least one CD, where $CD_\alpha = 2.4039$, $\alpha = 0.1$.

Figure 6 shows the CD diagram for accuracy. Specifically, all compared models whose average rank is within one CD of that of CCDS are connected. Otherwise, the model, which is not connected to CCDS, is perceived as having significantly different performance from the control approach. From Figure 6, we can see that the CCDS ranks 3rd among all the approaches, which performs significantly better than DMPWFC. However, compared with machine learning methods, it does not significantly outperform in terms of accuracy. However, ILMPFTC and S2CL significantly outperform SVM and NN in terms of accuracy. Thus, the proposed model (namely CCDS) can achieve highly competitive performance against the selected compared CCL methods, and has tied with the performance of the selected machine learning classification methods.

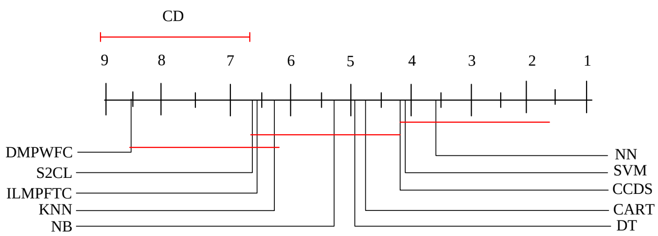


Figure 6. Comparison of the control model against other approaches with the Bonferroni-Dunn test ($CD = 2.4039$ at 0.1 significance level).

7. Conclusions

This paper discussed an attribute precedence concept-cognitive learning model based on knowledge space theory. Specifically, a δ-surmise function on attributes was first proposed. To provide enlightenment about the cognitive logic, the predecessor function, surmise degree, and cognitive error tolerance were introduced to develop the (δ,β)-surmise relation on attributes, by which a sparse adjacency relation was well-established to construct a granular-based cognitive transition mechanism. Then the cognitive transfer measure was naturally defined to make the significance of attributes available. Furthermore, considering the powerful node of concepts, such as the hub of communication, we also explored the discrimination of concepts to measure the nonconformity, which accompanied conceptual fusion. Moreover, an algorithm CCDS was designed and thoroughly performed on 20 datasets. Experimental results showed that CCDS's classification performance is better than that of the emerging mainstream CCL algorithms. In addition, the proposed model had no significant difference when compared with the traditional classification approaches of machine learning.

In summary, the proposed approach addresses an interesting and challenging field, providing new perspectives for CCL. Nevertheless, it needs to be further presented as a rationalized axiom system of the surmise system for fuzzy environments, which has not been further explored in this paper. The proposed model also fails to handle multi-decision scenarios. Consequently, the next tasks are to focus on the above crux matters.

Author contributions

First author (Ju Huang): Paper writing and establishment of specific models. Corresponding author (Yidong Lin): Overall research framework design. Last author (Wen Sun): Paper revision and provision of algorithmic support for improvement.

Use of AI tools declaration

The authors declare they have not used Artificial Intelligence (AI) tools in the creation of this article.

Acknowledgments

This work is supported by the National Natural Science Foundation of China (Nos. 12201284, 12301585), the Natural Science Foundation of Fujian Province, China (Nos. 2022J05169, 2023J05175, and 2024J01803), and the Fujian Province Teaching Reform Projects (No. FBJY20240024). Wei Yao made constructive suggestions for the revision of this work.

Conflict of interest

The authors declare no conflict of interest.

References

1. G. Birkhoff, Rings of sets, *Duke Math. J.*, **3** (1937), 443–454. https://doi.org/10.1215/S0012-7094-37-00334-X
2. L. Breiman, J. H. Friedman, R. A. Olshen, C. J. Stone, Classification and regression trees, *Wadsworth*, **40** (1984), 358–361. https://doi.org/10.2307/2530946

3. T. M. Cover, P. E. Hart, Nearest neighbor pattern classification, *IEEE T. Inform. Theory*, **13** (1967), 21–27. https://doi.org/10.1109/TIT.1967.1053964
4. D. X. Chen, J. J. Li, R. D. Lin, Formal concept analysis of multi-scale formal context, *J. Ambient Intell. Humaniz. Comput.*, **11** (2020), 5315–5327. https://doi.org/10.1007/s12652-020-01867-6
5. O. J. Dunn, Multiple comparisons among means, *J. Am. Stat. Assoc.*, **56** (1961), 52–64. https://doi.org/10.1080/01621459.1961.10482090
6. X. Y. Deng, J. H. Li, Y. H. Qian, J. M. Liu, Junmin, An emerging incremental fuzzy concept-cognitive learning model based on granular computing and conceptual knowledge clustering, *IEEE Trans. Emerg. Top. Comput. Intell.*, **8** (2024), 2417–2432. https://doi.org/10.1109/TETCI.2024.3360336
7. A. Elhadad, A. Ghareeb, S. Abbas, A blind and high-capacity data hiding of DICOM medical images based on fuzzification concepts, *Alex. Eng. J.*, **60** (2021), 2471–2482. https://doi.org/10.1016/j.aej.2020.12.050
8. J. C. Falmagne, J. P. Doignon, Learning spaces: interdisciplinary applied mathematics, *Springer-Verlag Berlin Heidelberg*, 2011. https://doi.org/10.1007/978-3-642-01039-2
9. M. Friedman, A comparison of alternative tests of significance for the problem of *m* rankings, *Annals Math. Stat.*, **11** (1940), 86–92. https://doi.org/10.1214/aoms/1177731944
10. N. Friedman, D. Geiger, M. Goldszmidt, Bayesian network classifiers, *Mach. Learn.*, **29** (1997), 131–163. https://doi.org/10.1023/A:1007465528199
11. B. J. Fan, E. C. C. Tsang, W. H. Xu, D. G. Cheng, W. T. Li, Attribute-oriented cognitive concept learning strategy: A multi-level method, *Int. J. Mach. Learn. Cybern.*, **10** (2019), 2421–2437. https://doi.org/10.1007/s13042-018-0879-5
12. M. Hu, E. C. C. Tsang, Y. T. Guo, Q. S. Zhang, D. G. Chen, W. H. Xu, A novel approach to concept-cognitive learning in interval-valued formal contexts: a granular computing viewpoint, *Int. J. Mach. Learn. Cybern.*, **13** (2022), 1049–1064. https://doi.org/10.1007/s13042-021-01434-1
13. J. H. Li, C. L. Mei, Y. J. Lv, Incomplete decision contexts: Approximate concept construction, rule acquisition and knowledge reduction, *Int. J. Approx. Reason.*, **54** (2013), 149–165. https://doi.org/10.1016/j.ijar.2012.07.005
14. J. H. Li, C. L. Mei, W. H. Xu, Y. H. Qian, Concept learning via granular computing: a cognitive viewpoint, *Inf. Sci.*, **298** (2015), 447–467. https://doi.org/10.1016/j.ins.2014.12.010
15. J. H. Li, C. C. Huang, J. J. Qi, Y. H. Qian, W. Q. Liu, Three-way cognitive concept learning via multi-granularity, *Inf. Sci.*, **378** (2017), 244–263. https://doi.org/10.1016/j.ins.2016.04.051
16. J. H. Li, Y. L. Mi, W. Q. Liu, Incremental cognition of concepts: theoreis and methods, *Chinese J. Comp.*, **42** (2019), 2233–2250. https://doi.org/10.11897/SP.J.1016.2019.02233
17. Z. M. Liu, J. H. Li, X. Zhang, X. Z. Wang, Incremental incomplete concept-cognitive learning model: A stochastic strategy, *IEEE T. Neur. Net. Lear.*, **36** (2025), 390–406. https://doi.org/10.1109/TNNLS.2023.3333537
18. Y. L. Mi, Concept-cognitive computing theory and model, *Beijing: University of Chinese Academy of Sciences*, 2018.
19. Y. L. Mi, J. H. Li, W. Q. Liu, J. Lin, Research on granular concept cognitive learning system under mapReduce framework, *Acta Electronica Sinica*, **46** (2018), 289–297. https://doi.org/10.3969/j.issn.0372-2112.2018.02.005

20. Y. L. Mi, Y. Shi, J. H. Li, W. Q. Liu, M. Y. Yan, Fuzzy-based concept learning method: Exploiting data with fuzzy conceptual clustering, *IEEE T. Cybernetics*, **52** (2022), 582–593. https://doi.org/10.1109/TCYB.2020.2980794
21. Y. L. Mi, W. Q. Liu, Y. Shi, J. H. Li, Semi-supervised concept learning by concept-cognitive learning and concept space, *IEEE T. Knowl. Data En.*, **34** (2020), 2429–2442. https://doi.org/10.1109/TKDE.2020.3010918
22. Y. L. Mi, Z. R. Wang, H. Liu, Y. Qu, G. F. Yu, Y. Shi, Divide and conquer: A granular concept-cognitive computing system for dynamic classification decision making, *Eur. J. Oper. Res.*, **308** (2023), 255–273. https://doi.org/10.1016/j.ejor.2022.12.018
23. Y. L. Mi, Concept neural network based on time-delay regret for dynamic stream learning, *IEEE T. Pattern Anal.*, **47** (2025), 3796–3814. https://doi.org/10.1109/TPAMI.2025.3535636
24. J. J. Qi, L. Wei, Y. Y. Yao, Three-way formal concept analysis, *Rough Sets and Knowledge Technology - 9th International Conference, Springer*, **8818** (2014), 732–741. https://doi.org/10.1007/978-3-319-11740-9_67
25. Y. Shi, Y. L. Mi, J. H. Li, W. Q. Liu, Concurrent concept-cognitive learning model for classification, *Inf. Sci.*, **496** (2019), 65–81. https://doi.org/10.1016/j.ins.2019.05.009
26. Y. Shi, Y. L. Mi, J. H. Li, W. Q. Liu, Concept-cognitive learning model for incremental concept learning, *IEEE T. Syst. Man Cy-S.*, **51** (2021), 809–821. https://doi.org/10.1109/TSMC.2018.2882090
27. E. C. C. Tsang, B. J. Fan, D. G. Chen, W. H. Xu, W. T. Li, Multi-level cognitive concept learning method oriented to data sets with fuzziness: a perspective from features, *Soft Comput.*, **24** (2020), 3753–3770. https://doi.org/10.1007/s00500-019-04144-7
28. R. Wille, Restructuring lattice theory: an approach based on hierarchies of concepts, *Restructuring lattice theory: an approach based on hierarchies of concepts, In Ordered Sets, I. Rival, Ed. Dordrecht, The Netherlands: Reidel*, **5584** (2009), 314–339. https://doi.org/10.1007/978-3-642-01815-2_23
29. I. H. Witten, E. Frank, M. A. Hall, Data mining: Practical machine learning tools and techniques, *SIGMOD Rec.*, **31** (2002), 76–77. https://doi.org/10.1145/507338.507355
30. L. D. Wang, X. D. Liu, Concept analysis via rough set and AFS algebra, *Inf. Sci.*, **178** (2008), 4125–4137. https://doi.org/10.1016/j.ins.2008.07.004
31. W. H. Xu, J. Z. Pang, S. Q. Luo, A novel cognitive system model and approach to transformation of information granule, *Int. J. Approx. Reason.*, **55** (2014), 853–866. https://doi.org/10.1016/j.ijar.2013.10.002
32. W. H. Xu, W. T. Li, Granular computing approach to two-way learning based on formal concept analysis in fuzzy datasets, *IEEE T. Cybernetics*, **46** (2016), 366–379. https://doi.org/10.1109/TCYB.2014.2361772
33. W. H. Xu, D. D. Guo, Y. H. Qian, W. P. Ding, Two-way concept-cognitive learning method: A fuzzy-based progressive learning, *IEEE T. Fuzzy Syst.*, **31** (2023), 1885–1899. https://doi.org/10.1109/TFUZZ.2022.3216110
34. W. H. Xu, Y. Q. Chen, Multi-attention concept-cognitive learning model: A perspective from conceptual clustering, *Knowl. Based Syst.*, **252** (2022), 109472. https://doi.org/10.1016/j.knosys.2022.109472

35. W. H. Xu, D. D. Guo, J. S. Mi, Y. H. Qian, K. Y. Zheng, W. P. Ding, Two-way concept-cognitive learning via concept movement viewpoint, *IEEE T. Neural Net. Lear.*, **34** (2023), 6798–6812. https://doi.org/10.1109/TNNLS.2023.3235800
36. X. X. Xie, W. H. Xu, J. J. Li, A novel concept-cognitive learning method: A perspective from competences, *Knowl. Based Syst.*, **265** (2023), 110382. https://doi.org/10.1016/j.knosys.2023.110382
37. S. Ben. Yahia, K. Arour, A. Slimani, A. Jaoua, Discovery of compact rules in relational databases, *Inf. Sci. J.*, **3** (2000), 497–511.
38. Y. Y. Yao, Interpreting concept learning in cognitive informatics and granular computing, *IEEE T. Syst. Man Cy. Part B*, **39** (2009), 855–866. https://doi.org/10.1109/TSMCB.2009.2013334
39. E. L. Yan, J. L. Song, C. F. Zheng, B. H. Mi, W. X. Hong, Construction of three-way attribute partial order structure via cognitive science and granular computing, *Knowl. Based Syst.*, **197** (2020), 105859. https://doi.org/10.1016/j.knosys.2020.105859
40. E. L. Yan, C. G. Yu, L. M. Lu, W. X. Hong, C. Z. Tang, Incremental concept cognitive learning based on three-way partial order structure, *Knowl. Based Syst.*, **220** (2021), 106898. https://doi.org/10.1016/j.knosys.2021.106898
41. K. H. Yuan, W. H. Xu, W. T. Li, W. P. Ding, An incremental learning mechanism for object classification based on progressive fuzzy three-way concept, *Inf. Sci.*, **584** (2022), 127–147. https://doi.org/10.1016/j.ins.2021.10.058
42. W. X. Zhang, W. H. Xu, Cognitive model based on granular computing, *Chinese J. Eng. Math.*, **24** (2007), 957–971. https://doi.org/10.1016/S1005-8885(07)60159-9
43. Y. X. Zhao, J. H. Li, W. Q. Liu, W. H. Xu, Cognitive concept learning from incomplete information, *Int. J. Mach. Learn. Cyb.*, **8** (2017), 159–170. https://doi.org/10.1007/s13042-016-0553-8
44. T. Zhang, H. H. Li, M. Q. Liu, M. Rong, Incremental concept-cognitive learning based on attribute topology, *Int. J. Approx. Reason.*, **118** (2020), 173–189. https://doi.org/10.1016/j.ijar.2019.12.010
45. T. Zhang, M. Rong, H. R. Shan, M. X. Liu, Stability analysis of incremental concept tree for concept cognitive learning, *Int. J. Mach. Learn. Cybern.*, **13** (2022), 11–28. https://doi.org/10.1007/s13042-021-01332-6
46. C. L. Zhang, E. C. C. Tsang, W. H. Xu, Y. D. Lin, L. Z. Yang, Incremental concept-cognitive learning approach for concept classification oriented to weighted fuzzy concepts, *Knowl. Based Syst.*, **260** (2023), 110093. https://doi.org/10.1016/j.knosys.2022.110093