Pattern Recognition 171 (2026) 112268

Contents lists available at ScienceDirect

Pattern Recognition

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

CM-CCL: Collaborative multi-scale concept-cognitive learning for knowledge discovery

Weihua Xu [a], Jinbo Wang [a,*], Qinghua Zhang [b,c]

[a] College of Artificial Intelligence, Southwest University, Chongqing, 400715, PR China
[b] School of Computer Science and Technology, Chongqing University of Posts and Telecommunications, Chongqing, 400065, PR China
[c] Key Laboratory of Cyberspace Big Data Intelligent Security, Ministry of Education, Chongqing, 400065, PR China

ARTICLE INFO

Keywords:
Concept-cognitive learning
Granular computing
Multi-scale data
Knowledge discovery

ABSTRACT

Concept-cognitive learning (CCL) provides an effective method for representing knowledge in data, with the use of concepts as knowledge carriers being its most significant characteristic. However, existing CCL models neglect the utilization of multi-scale information, resulting in insufficient representation capabilities of the learned concepts. Therefore, this paper proposes a novel multi-scale concept-cognitive learning model to address this issue. Firstly, a rational multi-scale data construction method is provided based on the characteristics of CCL. Then, a multi-scale feature selection method is introduced, which considers both the inter-scale correlations and intra-scale class distances. On this basis, progressive concepts are learned by integrating similar granular concepts at each scale to explicitly represent the knowledge in the data. Furthermore, a mechanism for the collaboration among progressive concepts at different scales is proposed to complete the classification task. Finally, a series of experiments are conducted to validate the effectiveness of the proposed CM-CCL model.

1. Introduction

With the advancement of information technology, massive and complex datasets data has been generated across various industries. Consequently, extracting valuable information from this massive dataset has become a common challenge. Some methods were proposed to reveal valuable information in data in the form of knowledge by simulating the way humans perceive things.

Knowledge presentation in data can take various forms, with concepts being a significant one [1]. Broadly speaking, concepts are the common characteristics of a group of entities [2]. They are derived through abstraction, allowing us to generalize the common features from a set of objects. Inspired by concepts in philosophy, Wille [3] utilized lattice theory to learn concepts from tabular data and represent knowledge in the form of concepts. The basic elements of concepts are the extent, which reflects information about objects, and the intent, which reflects information about features. All concepts learned from a dataset form a lattice structure, known as a concept lattice. Subsequently, some studies attempted to learn all concepts from the data and represent knowledge in the form of concept lattices. Early research focused on theoretical aspects, particularly on the construction of concept lattices [4]. These studies have laid the theoretical foundation for knowledge discovery with concepts as carriers. It is noted that, although concept lattices can fully reflect the relationships between objects and features, on the one hand, it is difficult to learn all the concepts from data, especially when dealing with large-scale datasets. On the other hand, the applicability of concept lattices in practical scenarios is relatively weak. Therefore, inspired by the patterns of human cognition, some research endeavored to learn concepts more efficiently, which leading to the development of the concept-cognitive learning (CCL) [5]. Notably, attempts to combine CCL with machine learning and granular computing theory have further enhanced its practicality [6]. In the theoretical domain, Xu et al. [7] explored the mechanism of information granulation transformation from clues to concepts, and they [8] further investigated concept learning method from movement viewpoint. The two-way concept learning models in fuzzy environment [9] were also studied. To enhance the representational power of concepts, a series of CCL methods were proposed, such as progressive concepts [10], three-way concepts [11], weighted concepts [12] and interval concepts [13]. In the application fields, Wu et al. [14] introduced a CCL approach that utilizes concepts for multi-label classification. Guo et al. [15] proposed a mechanism for the forgetting and recalling of concepts for dynamic knowledge fusion. Additionally, Wang et al. [16] presented a CCL model capable of processing multi-view data. These studies indicate that CCL can effectively represent and integrate the knowledge in data, using concepts as the basic carriers.

* Corresponding author.
E-mail addresses: chxuwh@gmail.com (W. Xu), wangjinbo2307@163.com (J. Wang), zhangqh@cqupt.edu.cn (Q. Zhang).

https://doi.org/10.1016/j.patcog.2025.112268
Received 10 March 2025; Received in revised form 4 July 2025; Accepted 9 August 2025
Available online 13 August 2025

In the process of human cognition, multi-scale features are consistently employed to achieve a comprehensive understanding of objects. This approach is crucial in fields such as computer vision, image processing and machine learning. Multi-scale features enable models to better understand objects by capturing information at different levels of data. Inspired by the human ability to process multi-scale information, some studies attempted to incorporate similar strategies to accomplish various tasks. For instance, deep neural networks are employed to extract multi-scale features from data, thereby enhancing the model's ability to understand complex patterns. This approach is widely applied in tasks such as change detection [17], image detection [18], and multi-task learning [19]. Wu and Leung [20] proposed a multi-scale data analysis model based on the idea of granular computing, in which objects can have multiple scale levels of values under each feature. This model is also referred to as the Wu-Leung model. On this basis, Li et al. [21] introduced a lattice-based generalized multi-scale data analysis model. Subsequently, based on the foundations laid in Wu and Leung [20], a series of methods were proposed to extract valuable information from multi-scale information systems. On the one hand, some studies tried to select the optimal scales from multi-scale information systems, based on the consistency between the binary relations formed by the features at different scales and the binary relations formed with decision information. Consequently, decision rules with improved generalization can be extracted at the optimal scale. For example, Yang et al. [22] proposed a method for selecting the optimal scale based on genetic algorithm. Wang et al. [23] investigated optimal scale selection for mixed data. Wang et al. [24] selected the optimal scale combination from multi-scale information systems and extracted decision rules based on multi-granulation rough set theory. On the other hand, there are also some studies that explored the use of multi-scale information to address knowledge discovery issues under various circumstances. For example, Zhang et al. [25] designed an unsupervised attribute reduction method by utilizing both multi-scale information and multiple correlations. Zhou et al. [26] proposed a semi-supervised feature selection method based on multi-scale fuzzy information. Wang et al. [27] performed feature selection by defining multi-scale fuzzy entropy. Zhang et al. [28] utilized multi-scale information to integrate multi-source data. Yin et al. [29] applied multi-scale information for multi-label learning. Yuan et al. [30] defined a novel zentropy uncertainty measure for multi-scale feature selection.

These studies demonstrate that multi-scale data play a significant role in knowledge discovery, which is mainly reflected in multi-scale data can provide richer measurement information. Compared with single-scale data, multi-scale data can characterize the relationships between objects and features at multiple scale levels. As a result, multi-scale data have stronger representation ability. By employing appropriate knowledge discovery methods, the knowledge hidden in the data can be effectively mined.

1.1. Motivation

Previous research on CCL has been conducted on tabular data at a single scale, meaning that an object is assigned a single value for a given feature. However, multi-scale data are more widely used in practice. For example, a student's grade can be precisely recorded on a percentage scale or roughly recorded as pass or fail. In fact, the ability to handle multi-scale information flexibly is a significant characteristic of human cognition. When recognizing an object, on the one hand, we need to grasp the overall characteristics from a coarser scale; on the other hand, we also need to learn some specific features from a finer-scale perspective. Intuitively, concept-cognitive learning, as a novel computational paradigm that simulates the process of human perception and understanding of objects, would significantly benefit from the incorporation of multi-scale information in the CCL process. Moreover, it should be noted that Wu et al. [20] have theoretically proposed a multi-scale data analysis model based on rough set theory, and numerous studies have been carried out on this basis. However, these studies have primarily focused on the theoretical selection of optimal scales while neglecting how to actively construct multi-scale data from single-scale data. Therefore, it is meaningful to explore multi-scale data construction methods applicable to CCL, starting with single-scale data. Based on the above analysis, in order to introduce multi-scale information into CCL models, this paper proposes a method for constructing multi-scale data, and further discusses how to learn and integrate concepts across multiple scales.

1.2. Contribution

Based on the above analysis, this paper proposes a novel collaborative multi-scale concept-cognitive learning model, and the overall framework of the proposed model is shown in Fig. 1. The main innovations are as follows:

(1) By considering the characteristics of CCL, a reasonable multi-scale data construction method is presented. It fully exploits the measurement information provided by each feature, thereby enhancing the representation ability of concepts at the data level.
(2) To enhance the efficiency of concept learning and eliminate redundant information, a multi-scale feature selection method is proposed that simultaneously considers intra-scale class distances and inter-scale correlations.
(3) By integrating similar concepts, a method for generating progressive concepts with enhanced knowledge representation capabilities from multi-scale data is studied. And the collaborative mechanism across progressive concepts at different scales is investigated to for classification.
(4) A multi-scale knowledge representation framework based on concepts is established, and its effectiveness is verified through a serious of experiments.

This paper is organized as follows. Section 2 reviews the basic notations relevant to this paper. Section 3 provides a detailed discussion of the proposed collaborative multi-scale concept-cognitive learning model. The experimental analysis is presented in Section 4, and Section 5 concluded this paper.

2. Preliminaries

In this section, foundational notations regarding fuzzy sets and fuzzy concepts are reviewed, which are discussed in detail in References [3] and [31].

2.1. Concepts

A dataset can be described as a triplet $F = (E, A, I)$, where $E = \{e_1, e_2, \dots, e_M\}$ represents the set of objects, $A = \{a_1, a_2, \dots, a_N\}$ denotes the set of features, and $I : E \times A \to \{0, 1\}$ is a Boolean relation on the Cartesian product $E \times A$. F is also called a formal context in formal concept analysis theory. Here, for any $e \in E$ and $a \in A$, $I(e, a) = 1$ indicates that object e possesses feature a, while $I(e, a) = 0$ represents that object e does not possess feature a.

To represent the knowledge in the data in the form of concept, a pair of operators $f : 2^E \to 2^A$ and $g : 2^A \to 2^E$ can be defined [3]. More specifically, for any $X \in 2^E$ and $B \in 2^A$,

$$\begin{aligned} f(X) &= \{a \in A | I(e, a) = 1, \forall e \in X\}, \\ g(B) &= \{e \in E | I(e, a) = 1, \forall a \in B\}. \end{aligned} \tag{1}$$

2^E and 2^A are, respectively, the power set of E and A. Then, a pair (X, B) is called a concept if and only if $f(X) = B$ and $g(B) = X$. X and B are, respectively, referred to as the extent and intent of (X, B). Intuitively, X is the set of objects that possess all the features in B, and B is the shared feature set that possessed by the objects in X. Concepts have the following properties [3].

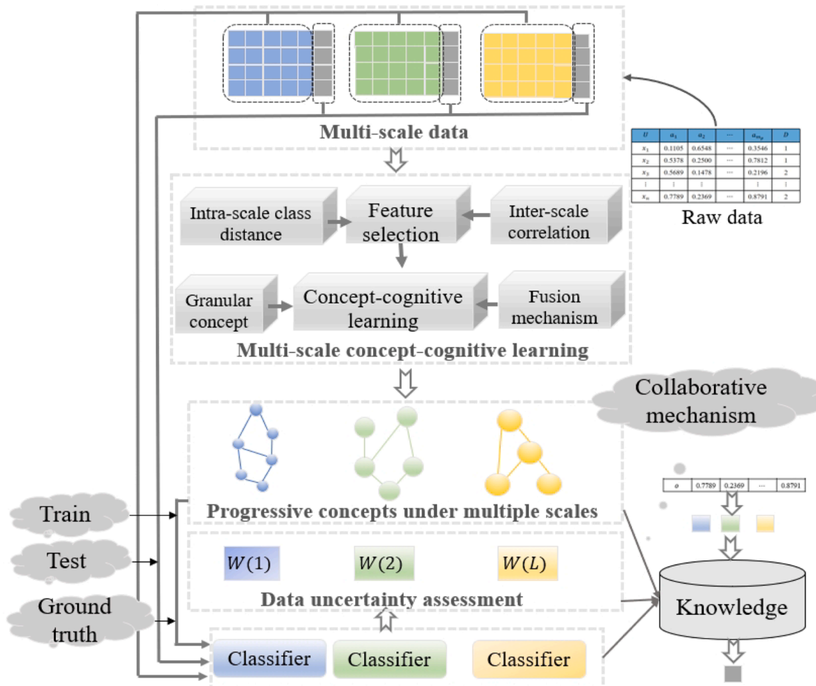

Fig. 1. Procedure of collaborative multi-scale concept-cognitive learning model.

Property 1. *For $X_1, X_2 \in 2^E$ and $B_1, B_2 \in 2^A$, we have:*

(1) $X_1 \subseteq X_2 \Rightarrow f(X_2) \subseteq f(X_1)$, $B_1 \subseteq B_2 \Rightarrow g(B_2) \subseteq g(B_1)$;
(2) $fgf(X_1) = f(X_1)$, $gfg(B_1) = g(B_1)$;
(3) $(gf(X_1), f(X_1))$ *and* $(g(B_1), fg(B_1))$ *are concepts.*

In Property 1, $f(X_1) \in 2^A$, $gf(X_1) \in 2^E$, and $fgf(X_1) \in 2^A$. According to (3), we know that starting from the given learning clues (which can be a set of objects or a set of features), the concept can be learned after two mappings through the operators f and g.

A dataset containing label information can be further represented as a quintuple $F = (E, A, I, D, J)$, where $D = \{d_1, d_2, \dots, d_K\}$ is the set of labels, and $J : E \times D \to \{0, 1\}$ is a Boolean relation on $E \times D$. Specifically, for $e \in E$ and $d_k \in D$, $J(e, d_k) = 1$ denotes that object e possesses label d_k, while $J(e, d_k) = 0$ indicates that object e does not possess label d_k. In this paper, it is assumed that each object has and only has one label. Therefore, the objects in E form a partition $\{D_1, D_2, \dots, D_K\}$ with respect to the label set D, where for each $k \in \{1, 2, \dots, K\}$), $D_k = \{e \in E | J(e, d_k) = 1\}$.

In order to learn the concepts associated with a given label, some studies attempt to learn concept from a local perspective. Given a dataset $F = (E, A, I, D, J)$, for any $d_k \in D$, $X \in 2^{D_k}$ and $B \in 2^A$, a pair of local operators $f^k : 2^{D_k} \to 2^A$ and $g^k : 2^A \to 2^{D_k}$ is defined as follows:

$$\begin{aligned} f_k(X) &= \{a \in A | I(e, a) = 1, \forall e \in D_k\}, \\ g_k(B) &= \{e \in D_k | I(e, a) = 1, \forall a \in B\}. \end{aligned} \tag{2}$$

Based on f_k and g_k, the concepts that associated with the label d_k can be learned.

2.2. Fuzzy concepts

In the dataset defined in Section 2.1, an object either possesses a feature or does not possess a feature. However, in practical applications, an object may possess a feature to a certain degree. Therefore, the definition of fuzzy concepts was developed to learn concepts from such fuzzy data.

Given a feature set $A = \{a_1, a_2, \dots, a_N\}$, the collection of all fuzzy sets on A is denoted as $\mathcal{F}^A$. For any $\widetilde{B} \in \mathcal{F}^A$, $\widetilde{B} = \{\frac{\widetilde{B}(a_1)}{a_1}, \frac{\widetilde{B}(a_2)}{a_2}, \dots, \frac{\widetilde{B}(a_N)}{a_N}\}$, where $\widetilde{B}(a_n) \in [0, 1]$ ($n \in \{1, 2, \dots, N\}$) is degree of membership of a_n in $\widetilde{B}$. It can also be understood as the degree to which $\widetilde{B}$ possesses the feature a_n. For $\widetilde{B}_1, \widetilde{B}_2 \in \mathcal{F}^A$, the inclusion relationship $\subseteq$ between $\widetilde{B}_1$ and $\widetilde{B}_2$ is defined as follows:

$$\widetilde{B}_1 \subseteq \widetilde{B}_2 \Leftrightarrow \widetilde{B}_1(a) \leq \widetilde{B}_2(a), \forall a \in A. \tag{3}$$

A fuzzy dataset can be represented as a quintuple $F = (E, A, \widetilde{I}, D, J)$. The definitions of E, A, D, and J are the same as that in Section 2.1. $\widetilde{I} : E \times A \to [0, 1]$ is a fuzzy binary relation on $E \times A$. For $e \in E$ and $a \in A$, $\widetilde{I}(e, a)$ represents the degree to which e possesses the feature a, it can also be understood as the value of object e under feature a.

To learn concepts from fuzzy datasets, a pair of local operators $\widetilde{f_k}$ and $\widetilde{g_k}$ can be defined [4,31]. For $d_k \in D$, $\widetilde{B} \in \mathcal{F}^A$ and $X \in 2^{D_k}$,

$$\begin{aligned} \widetilde{f_k}(X)(a) &= \bigwedge_{e \in X} \widetilde{I}(e, a), a \in A, \\ \widetilde{g_k}(\widetilde{B}) &= \{e \in D_k | \widetilde{B}(a) \leq \widetilde{I}(e, a), \forall a \in A\}. \end{aligned} \tag{4}$$

The symbol $\bigwedge$ denotes the minimum operation. Specifically, given a feature a, $\bigwedge_{e \in X} \widetilde{I}(e, a)$ denotes the minimum value in the set $\{\widetilde{I}(e, a) | e \in X\}$. It should be noticed that $\widetilde{g_k}(\widetilde{B})$ is a crisp set, and $\widetilde{f_k}(X)$ is a fuzzy set where $\widetilde{f_k}(X)(a)$ represent the degree to which $\widetilde{f_k}(X)$ possesses a. Based on $\widetilde{f_k}$ and $\widetilde{g_k}$, a type of fuzzy concept can be learned. For $X \in 2^{D_k}$ and $\widetilde{B} \in \mathcal{F}^A$, a pair $(X, \widetilde{B})$ is referred to as a fuzzy concept if

$$\widetilde{f_k}(X) = \widetilde{B}, \widetilde{g_k}(\widetilde{B}) = X. \tag{5}$$

Let $\mathcal{L}_k$ be the set of all fuzzy concepts learned from the data associated with label d_k. A partial order $\preceq$ can be defined on $\mathcal{L}_k$. For any $(X_1, \widetilde{B}_1), (X_2, \widetilde{B}_2) \in \mathcal{L}_k$,

$$(X_1, \widetilde{B}_1) \preceq (X_2, \widetilde{B}_2) \Leftrightarrow X_1 \subseteq X_2 \Leftrightarrow \widetilde{B}_2 \subseteq \widetilde{B}_1. \tag{6}$$

According to the above formula, there is a generalization-specialization relationship between $(X_1, \widetilde{B}_1)$ and $(X_2, \widetilde{B}_2)$, where $(X_1, \widetilde{B}_1)$ is called the specialization concept of $(X_2, \widetilde{B}_2)$, and $(X_2, \widetilde{B}_2)$ is called the generalization concept of $(X_1, \widetilde{B}_1)$. $(X_1, \widetilde{B}_1)$ is also called a sub-concept of $(X_2, \widetilde{B}_2)$ Intuitively, the more objects a fuzzy concept's extent contains, the fewer common features they share.

When the values of objects under features degenerate from the interval $[0, 1]$ to the set $\{0, 1\}$, the fuzzy concepts defined in this section degenerate into the concepts described in Section 2.1. Therefore, the application range of fuzzy concepts is broader. In the subsequent discussion, we will refer to fuzzy concepts simply as concepts without causing confusion. Moreover, as fuzzy data is more frequently encountered in real-world scenarios, the datasets discussed subsequently are fuzzy datasets.

3. Collaborative multi-scale concept-cognitive learning

Human cognition is inherently multi-scale, and concept-cognitive learning, as a novel computational paradigm that simulates human perception and understanding of objects, would significantly benefit from the incorporation of multi-scale information. Thus, this section attempts to introduce multi-scale information into the concept learning process, according to the characteristics of CCL.

3.1. Construction of multi-scale data

By combining the characteristics of CCL, the definition of how to generate multi-scale data is presented in Definition 1.

Definition 1. Let $F=(E,A,\widetilde{I},D,J)$ be a dataset, where $E=\{e_1,e_2,\dots,e_M\}$, $A=\{a_1,a_2,\dots,a_N\}$, and $D=\{d_1,d_2,\dots,d_K\}$. The multi-scale dataset obtained from F is defined as $MF=\{F_1,F_2,\dots,F_L\}$, where:

(1) L is the parameter representing the number of scales, which takes on positive integer values;
(2) $F_l=(E,A_l,\widetilde{I}_l,D,J)$ is the dataset under the lth scale, where $A_l=\{a_n^l|n=1,2,\dots,N\}$. The notation a_n^l represents the feature of a_n at the lth scale. $\widetilde{I}_l: E\times A_l\to[0,1]$ is the fuzzy binary relation on $E\times A_l$, where $\widetilde{I}_l(e,a_n^l)$ represents the degree to which object e possesses feature a_n^l;
(3) For any $e\in E$ and $a_n\in A$,

$$\widetilde{I}_l(e,a_n^l)=b+(c-b)\frac{\left(\widetilde{I}(e,a_n)\right)^l-b^l}{c^l-b^l},\tag{7}$$

where $(\cdot)^l$ represents the lth power, and c and b are the maximum and minimum values of all objects in E under feature a_n, respectively.

Since the label information is already encoded in a_n^l, for simplicity we use $\widetilde{I}(e,a_n^l)$ to denote $\widetilde{I}_l(e,a_n^l)$ in the following discussion. It should be noticed that the set of objects E, the set of labels D, and the binary relation J are the same across all scales. The algorithm to generate multi-scale datasets is given in Algorithm 1. If we apply the local operators in Eq. (4) on F_l, concepts under the lth scale can be learned. For $X\in 2^{D_k}$ and $\widetilde{B}\in\mathcal{F}^{A_l}$, the local operators are denoted as $\widetilde{f}_k^l$ and $\widetilde{g}_k^l$, where

$$\begin{aligned}&\widetilde{f}_k^l(X)(a_n^l)=\bigwedge_{e\in X}\widetilde{I}(e,a_n^l),a_n^l\in A_l,\\&\widetilde{g}_k^l(\widetilde{B})=\{e\in D_k|\widetilde{B}(a_n^l)\le\widetilde{I}(e,a_n^l),\forall a_n^l\in A_l\}.\end{aligned}\tag{8}$$

Algorithm 1: Generating multi-scale datasets.

Input: A dataset $F=(E,A,\widetilde{I},D,J)$, and the number of scales L.
Output: The generated multi-scale dataset $MF=\{F_l,F_2,\dots,F_L\}$.
for $l\in\{1,2,\dots,L\}$ **do**
  for $a_n\in\{a_1,a_2,\dots,a_N\}$ **do**
    for $e\in E$ **do**
      Obtain $\widetilde{I}_l(e,a_n^l)$ according to Eq. (7);
  Obtain the dataset under the lth scale $F_l=(E,A_l,\widetilde{I}_l,D,J)$;
return the generated multi-scale dataset $MF=\{F_1,F_2,\dots,F_L\}$.

Compared to F, MF contains richer measurement information and has some special multi-scale properties.

Property 2. *Let $MF=\{F_1,F_2,\dots,F_L\}$ be a multi-scale dataset, we have:*

(1) For any $e_{m_1},e_{m_2}\in E$, $a_n\in A$, and $l\in\{1,2,\dots,L\}$, if $\widetilde{I}(e_{m_1},a_n)\ge\widetilde{I}(e_{m_2},a_n)$, then
$\widetilde{I}(e_{m_1},a_n^l)\ge\widetilde{I}(e_{m_2},a_n^l)$;
(2) For any $e\in E$, $a_n\in A$, and $l_1,l_2\in\{1,2,\dots,L\}$, if $l_2\ge l_1$, then
$\widetilde{I}(e,a_n^{l_1})\ge\widetilde{I}(e,a_n^{l_2})$;
(3) For $X\in 2^{D_k}$ and $l_1,l_2\in\{1,2,\dots,L\}$, $(\widetilde{g}_k^{l_1}\widetilde{f}_k^{l_1}(X),\widetilde{f}_k^{l_1}(X))$ and $(\widetilde{g}_k^{l_2}\widetilde{f}_k^{l_2}(X),\widetilde{f}_k^{l_2}(X))$ are concepts, and
$\widetilde{g}_k^{l_1}\widetilde{f}_k^{l_1}(X)=\widetilde{g}_k^{l_2}\widetilde{f}_k^{l_2}(X)$.

Proof.

(1) It is obvious according to Eq. (7).
(2) Since $l_2\ge l_1$, we have $\left(\widetilde{I}(e,a_n)\right)^{l_l}\ge\left(\widetilde{I}(e,a_n)\right)^{l_2}$. Then, according to Eq. (7), $\widetilde{I}(e,a_n^{l_1})\ge\widetilde{I}(e,a_n^{l_2})$.
(3) According to Reference [4], $(\widetilde{g}_k^{l_1}\widetilde{f}_k^{l_1}(X),\widetilde{f}_k^{l_1}(X))$ and $(\widetilde{g}_k^{l_2}\widetilde{f}_k^{l_2}(X),\widetilde{f}_k^{l_2}(X))$ are concepts that can be proven.

We now prove $\widetilde{g}_k^{l_1}\widetilde{f}_k^{l_1}(X)=\widetilde{g}_k^{l_2}\widetilde{f}_k^{l_2}(X)$ as follows. For any $e_{m_1},e_{m_2}\in D_k$ and $a_n\in A$, according to Eq. (7), we get

$$\widetilde{I}(e_{m_2},a_n)\le\widetilde{I}(e_{m_1},a_n)\Leftrightarrow\widetilde{I}(e_{m_2},a_n^{l_1})\le\widetilde{I}(e_{m_1},a_n^{l_1}),\tag{9}$$

$$\widetilde{I}(e_{m_2},a_n)\le\widetilde{I}(e_{m_1},a_n)\Leftrightarrow\widetilde{I}(e_{m_2},a_n^{l_2})\le\widetilde{I}(e_{m_1},a_n^{l_2}).\tag{10}$$

It implies that

$$\widetilde{I}(e_{m_2},a_n^{l_1})\le\widetilde{I}(e_{m_1},a_n^{l_1})\Leftrightarrow\widetilde{I}(e_{m_2},a_n^{l_2})\le\widetilde{I}(e_{m_1},a_n^{l_2}).\tag{11}$$

Based on Eq. (11), for any $X\in 2^{D_k}$, we know that

$$\bigwedge_{e\in X}\widetilde{I}(e,a_n^{l_1})\le\widetilde{I}(e_{m_1},a_n^{l_1})\Leftrightarrow\bigwedge_{e\in X}\widetilde{I}(e,a_n^{l_2})\le\widetilde{I}(e_{m_1},a_n^{l_2}).\tag{12}$$

Thus,

$$\widetilde{f}_k^{l_1}(X)(a_n^{l_1})\le\widetilde{I}(e_{m_1},a_n^{l_1})\Leftrightarrow\widetilde{f}_k^{l_2}(X)(a_n^{l_2})\le\widetilde{I}(e_{m_1},a_n^{l_2}).\tag{13}$$

Based on Eqs. (8) and (13), it follows that $\widetilde{g}_k^{l_1}\widetilde{f}_k^{l_1}(X)=\widetilde{g}_k^{l_2}\widetilde{f}_k^{l_2}(X)$.

□

Property 2 demonstrates that: (1) The multi-scale data does not alter the relative degree of features possession among objects within the original data; (2) Starting from the same clue X, the extent of the learned concepts under different scales are the same. An example is given in Example 1 to illustrate this progress.

Example 1. A dataset F is represented in Table 1, where $E=\{e_1,e_2,e_3,e_4,e_5,e_6,e_7\}$, $A=\{a_1,a_2,a_3,a_4\}$ and $D=\{d_1,d_2\}$. Suppose $L=3$, then the generated multi-scale dataset is $MF=\{F_1,F_2,F_3\}$. The datasets at the first, second, and third scales are presented in Tables 2–4, respectively. Let $X=\{e_5,e_6\}\subseteq D_2$, based on the discussion in this section, it can be calculated that:

$$(\widetilde{g}_2^1\widetilde{f}_2^1(X),\widetilde{f}_2^1(X))=\left(\{e_5,e_6\},\left\{\frac{0.70}{a_1^1},\frac{0.10}{a_2^1},\frac{0.10}{a_3^1},\frac{0.80}{a_4^1}\right\}\right),$$

$$(\widetilde{g}_2^2\widetilde{f}_2^2(X),\widetilde{f}_2^2(X))=\left(\{e_5,e_6\},\left\{\frac{0.58}{a_1^2},\frac{0.10}{a_2^2},\frac{0.10}{a_3^2},\frac{0.73}{a_4^2}\right\}\right),$$

$$(\widetilde{g}_2^3\widetilde{f}_2^3(X),\widetilde{f}_2^3(X))=\left(\{e_5,e_6\},\left\{\frac{0.48}{a_1^3},\frac{0.10}{a_2^3},\frac{0.10}{a_3^3},\frac{0.66}{a_4^3}\right\}\right).$$

Notice that at different scales, based on the same clue, one can learn concepts that have the same extent but different intent.

In Example 1, as the scale increases, it can be observed that objects with a low degree of possession of a certain feature will become even less possessive of that feature, while objects with a higher degree of possession will relatively become more possessive. In other words, it can be understood that as the scale increases, some noise will be gradually eliminated, and relatively important information will be retained.

Table 1
An example of a dataset.

Objects	a_1	a_2	a_3	a_4	d_1	d_2
e_1	0.10	0.80	0.70	0.15	1	0
e_2	0.10	0.90	0.90	0.10	1	0
e_3	0.30	0.90	0.85	0.20	1	0
e_4	0.50	0.30	0.20	0.70	0	1
e_5	0.70	0.10	0.15	0.80	0	1
e_6	0.90	0.20	0.10	0.90	0	1
e_7	0.70	0.10	0.10	0.70	0	1

Table 2
The first scale of the multi-scale dataset.

Objects	a_1^1	a_2^1	a_3^1	a_4^1	d_1	d_2
e_1	0.10	0.80	0.70	0.15	1	0
e_2	0.10	0.90	0.90	0.10	1	0
e_3	0.30	0.90	0.85	0.20	1	0
e_4	0.50	0.30	0.20	0.70	0	1
e_5	0.70	0.10	0.15	0.80	0	1
e_6	0.90	0.20	0.10	0.90	0	1
e_7	0.70	0.10	0.10	0.70	0	1

Table 3
The second scale of the multi-scale dataset.

Objects	a_1^2	a_2^2	a_3^2	a_4^2	d_1	d_2
e_1	0.10	0.73	0.58	0.11	1	0
e_2	0.10	0.90	0.90	0.10	1	0
e_3	0.18	0.90	0.82	0.13	1	0
e_4	0.34	0.18	0.13	0.58	0	1
e_5	0.58	0.10	0.11	0.73	0	1
e_6	0.90	0.13	0.10	0.90	0	1
e_7	0.58	0.10	0.10	0.58	0	1

Table 4
The third scale of the multi-scale dataset.

Objects	a_1^3	a_2^3	a_3^3	a_4^3	d_1	d_2
e_1	0.10	0.67	0.48	0.10	1	0
e_2	0.10	0.90	0.90	0.10	1	0
e_3	0.13	0.90	0.77	0.11	1	0
e_4	0.24	0.13	0.10	0.48	0	1
e_5	0.48	0.10	0.10	0.66	0	1
e_6	0.90	0.10	0.10	0.90	0	1
e_7	0.48	0.10	0.10	0.48	0	1

3.2. Feature selection based on inter-scale correlation and intra-scale class distance

In multi-scale datasets, the rich information provided by features are mined. However, some generated features may have weak representational power or be redundant. Therefore, by considering the characteristic of CCL, a feature selection method that simultaneously considers inter-scale correlations and intra-scale class distances is proposed in this subsection.

For a multi-scale dataset $MF = \{F_1, F_2, \dots, F_L\}$, let $AT = \{a_n^l | n = 1, 2, \dots, N, l = 1, 2, \dots, L\}$ be the set of features under all scales. First, the following definitions are given to measure the intra-scale class distances.

Definition 2. For any $a_n^l \in AT$, the distance between $e_{m_1}, e_{m_2} \in E$ under a_n^l is define as follows:

$$dis(a_n^l | e_{m_1}, e_{m_2}) = \begin{cases} 0, & J(e_{m_1}, :) = J(e_{m_2}, :), \\ ||\widetilde{I}(e_{m_1}, a_n^l) - \widetilde{I}(e_{m_2}, a_n^l)||_1, & J(e_{m_1}, :) \neq J(e_{m_2}, :). \end{cases} \tag{14}$$

$||\cdot||_1$ represents 1-norm. $J(e_{m_1}, :) = J(e_{m_2}, :)$ indicates that e_{m_1} and e_{m_2} have the same label, i.e., for all $d_k \in D$, $J(e_{m_1}, d_k) = J(e_{m_2}, d_k)$. $J(e_{m_1}, :) \neq J(e_{m_2}, :)$ represents that there exists $d_k \in D$ such that $J(e_{m_1}, d_k) \neq J(e_{m_2}, d_k)$.

Based on Definition 2, the class distance under each feature is defined as follows.

Definition 3. For any $a_n^l \in AT$, the class distance with respect to a_n^l is defined as follows:

$$cd(a_n^l) = \sum_{e_{m_1} \in E} \sum_{e_{m_2} \in E} dis(a_n^l | e_{m_1}, e_{m_2}). \tag{15}$$

$cd(a_n^l)$ reflects the discriminative power of feature a_n^l in distinguishing objects across different classes.

Then, the Pearson correlation coefficient is used to measure the correlation among all features in AT. For any $a_{n_1}^{l_1}, a_{n_2}^{l_2} \in AT$, the Pearson correlation coefficient $P(a_{n_1}^{l_1}, a_{n_2}^{l_2})$ between them is defined as follows,

$$P(a_{n_1}^{l_1}, a_{n_2}^{l_2}) = \frac{\sum_{e \in E}(\widetilde{I}(e, a_{n_1}^{l_1}) - \overline{a_{n_1}^{l_1}})(\widetilde{I}(e, a_{n_2}^{l_2}) - \overline{a_{n_2}^{l_2}})}{\sqrt{\sum_{e \in E}(\widetilde{I}(e, a_{n_1}^{l_1}) - \overline{a_{n_1}^{l_1}})^2 \sum_{e \in E}(\widetilde{I}(e, a_{n_2}^{l_2}) - \overline{a_{n_2}^{l_2}})^2}}, \tag{16}$$

where $\overline{a_{n_1}^{l_1}}$ represents the average value of feature $a_{n_1}^{l_1}$, and $\overline{a_{n_2}^{l_2}}$ is the average value of $a_{n_2}^{l_2}$.

Then, the features that are both representative and discriminative are expected to be selected from AT. Discriminability is given by Eq. (15), while the Pearson correlation coefficient can reflect representativeness to some extent. Based on this idea, the proposed feature selection method can paper be divided into the following two steps.

(1) Calculate the cd value of each feature in AT according to Definition 3, and sort all them from largest to smallest based on the cd values. On the basis of the sorted features, calculate the correlation coefficient matrix P among these features.

(2) Based on P, we aim to select a feature subset $SAT = \{a_{n_1}^{l_1}, a_{n_2}^{l_2}, \dots, a_{n_i}^{l_i}\}$ from AT, where the features in SAT correspond to the top-i features with the largest cd values in AT $(cd(a_{n_1}^{l_1}) \geq cd(a_{n_2}^{l_2}) \geq \dots \geq cd(a_{n_i}^{l_i}))$. First, we define the set of features whose Pearson correlation coefficient with a_n^l exceeds δ as $corr_{a_n^l}^{\delta} = \{a_{n_j}^{l_j} | P(a_n^l, a_{n_j}^{l_j}) > \delta\}$. The features in $corr_{a_n^l}^{\delta}$ can be represented by a_n^l within the threshold δ. Then, we aim to select the top-i features that satisfy the following two conditions:

$$corr_{a_{n_1}^{l_1}}^{\delta} \bigcup corr_{a_{n_2}^{l_2}}^{\delta} \bigcup \dots \bigcup corr_{a_{n_{i-1}}^{l_{i-1}}}^{\delta} \subset AT, \tag{17}$$

$$corr_{a_{n_1}^{l_1}}^{\delta} \bigcup corr_{a_{n_2}^{l_2}}^{\delta} \bigcup \dots \bigcup corr_{a_{n_i}^{l_i}}^{\delta} = AT, \tag{18}$$

where $\delta \in [-1, 1]$ is a parameter.

After obtaining SAT, retaining the features existing in SAT at each scale can obtain the multi-scale dataset after feature selection. The purpose of the above process is to identify features that are both discriminative and representative at each scale, thereby enhancing the representation ability of the subsequently learned concepts from a data-driven perspective. The detailed method to obtain SAT is given in Algorithm 2. Based on Algorithm 2, the features with strong discriminability and representativeness are selected at each scale.

Example 2. Continued from Example 1. If $\delta = 0.5$, according to Algorithm 3, it can be calculated that $A_1' = \{a_2^1, a_3^1, a_4^1\}$, $A_2' = \{a_2^2, a_3^2\}$, $A_3' = \{a_2^3\}$. If $\delta = 1$, then $A_1' = \{a_1^1, a_2^1, a_3^1, a_4^1\}$, $A_2' = \{a_1^2, a_2^2, a_3^2, a_4^2\}$, $A_3' = \{a_1^3, a_2^3, a_3^3, a_4^3\}$. After feature selection, the corresponding multi-scale dataset $MF' = \{F_1', F_2', F_3'\}$ is also obtained, where $F_1' = (E, A_1', \widetilde{I}_1, D, J)$, $F_2' = (E, A_2', \widetilde{I}_2, D, J)$, and $F_3' = (E, A_3', \widetilde{I}_3, D, J)$.

Algorithm 2: Feature selection in multi-scale datasets.

Input: A multi-scale dataset $MF = \{F_1, F_2, \dots, F_L\}$, the parameter δ.
Output: The multi-scale dataset $MF' = \{F'_l | l = 1, 2, \dots, L\}$ after feature selection.
Calculate the cd value of all feature in AT and sort them in descending order;
Calculate the Pearson correlation coefficients matrix P;
for $i \in \{2, 3, \dots, |AT|\}$ **do**
  $Temp_1 \leftarrow corr^{\delta}_{a^{l_1}_{n_1}} \bigcup corr^{\delta}_{a^{l_2}_{n_2}} \bigcup \dots \bigcup corr^{\delta}_{a^{l_{i-1}}_{n_{i-1}}}$;
  $Temp_2 \leftarrow corr^{\delta}_{a^{l_1}_{n_1}} \bigcup corr^{\delta}_{a^{l_2}_{n_2}} \bigcup \dots \bigcup corr^{\delta}_{a^{l_i}_{n_i}}$;
  if $Temp_1 \subset AT$ *and* $Temp_2 = AT$ **then**
    break;
SAT are top-i features with the largest cd values in AT;
for $l \in \{1, 2, \dots, L\}$ **do**
  $A'_l \leftarrow A_l \cap SAT$;
return the multi-scale dataset $MF' = \{F'_1, F'_2, \dots, F'_L\}$ after feature selection.

3.3. Concept-cognitive learning for multi-scale data

In this subsection, the CCL model for the multi-scale dataset $MF' = \{F'_1, F'_2, \dots, F'_L\}$ after feature selection is investigated.

In MF', for any $d_k \in D$, $X \in 2^{D_k}$ and $\widetilde{B} \in \mathcal{F}^{A'_l}$, the local operators $\widetilde{f^l_k}$ and $\widetilde{g^l_k}$ to learn concepts at the lth scale associated with label d_k are defined as follows,

$$\widetilde{f^l_k}(X)(a^l_n) = \bigwedge_{e \in X} \widetilde{I}(e, a^l_n), a^l_n \in A'_l, \qquad \widetilde{g^l_k}(\widetilde{B}) = \{e \in D_k | \widetilde{B}(a^l_n) \le \widetilde{I}(e, a^l_n), \forall a^l_n \in A'_l\}. \tag{19}$$

Based on the discussion in Section 2, it is extremely difficult to learn all concepts from a dataset. Therefore, granular concepts learned from individual objects as clues received considerable attention. For example, $(\widetilde{g^l_k}\widetilde{f^l_k}(\{e\}), \widetilde{f^l_k}(\{e\}))$ is a concept induced by $e \in D_k$ under the lth scale, and it is simply recorded as $(\widetilde{g^l_k}\widetilde{f^l_k}(e), \widetilde{f^l_k}(e))$. Let $G^l_k = \{(\widetilde{g^l_k}\widetilde{f^l_k}(e), \widetilde{f^l_k}(e)) | e \in D_k\}$ is be set of granular concepts with respect to D_k under the lth scale.

It should be noted that while granular concepts are relatively easy to learn, they possess a weaker representational capability. This implies that the intents of granular concepts may not adequately reflect the characteristics of the features associated with d_k. Therefore, some methods attempt to obtain concepts with stronger representational capabilities. Inspired by Yuan et al. [10] and Mi et al. [32], this subsection proposes a CCL method that integrates similar granular concepts to enhance their representational power. A concept $(X, \widetilde{B}) \in G^l_k$ is called a maximal concept in G^l_k, if:

(1) There exists $(X_1, \widetilde{B}_1) \in G^l_k$ such that $X_1 \subseteq X$;
(2) There does not exist $(X_2, \widetilde{B}_2) \in G^l_k$ such that $X \subset X_2$.

Based on maximal concepts, the method to learn progressive concepts is defined in the following way.

Definition 4. Let $(X, \widetilde{B})$ be a maximal concept in G^l_k, suppose $\{(X_i, \widetilde{B}_i) | i = 1, 2, \dots, j\} \subseteq G^l_k$ and $X_i \subseteq X$ $(i \in \{1, 2, \dots, j\})$. A progressive concept $(\mathcal{X}, \mathcal{B})$ is defined as follows:

$$\mathcal{X} = X_1 \cup X_2 \cup X_3 \cup \dots \cup X_j, \qquad \mathcal{B} = \frac{1}{2^{j-1}}(\widetilde{B}_1 + \widetilde{B}_2 + 2\widetilde{B}_3 + \dots + 2^{j-2}\widetilde{B}_j), \tag{20}$$

where $|X_1| \le |X_2| \le \dots \le |X_j|$.

It can be seen that progressive concepts are formed by integrating maximal concepts with their sub-concepts. The more objects a concept's extent encompasses, the more effectively it reflects the core characteristics of those objects. Consequently, during the fusion process, its corresponding weight is increased.

The set of progressive concepts learned from G^l_kn is called a concept subspace associated with label d_k under the lth scale, and is denoted as CS^l_k. The detailed process of generating CS^l_k is elaborated in Algorithm 3. Based on CS^l_k, a concept space $CS = \{CS^l_k | k = 1, 2, \dots, K, l = 1, 2, \dots, L\}$ is then generated.

Algorithm 3: The construction of progressive concepts.

Input: A set of granular concept G^l_k.
Output: The concept subspace CS^l_k.
Let $\{(X_i, \widetilde{B}_i) | i = 1, 2, \dots, p\}$ be the set maximal concepts in G^l_k, and $|X_1| \ge |X_2| \ge \dots \ge |X_p|$;
for $i \in \{1, 2, \dots, p\}$ **do**
  Find the sub-concepts set S of $(X_i, \widetilde{B}_i)$ from G^l_k;
  Compute $(\mathcal{X}, \mathcal{B})$ according to Definition 4;
  Add $(\mathcal{X}, \mathcal{B})$ to CS^l_k;
  $G^l_k \leftarrow G^l_k - S$;
return CS^l_k.

Example 3. Continued from Example 2, if $\delta = 0.5$, it can be calculated that:

$$CS^1_1 = \Big\{ \big(\{e_1, e_3\}, \{\frac{0.78}{a^1_2}, \frac{0.10}{a^1_3}, \frac{0.18}{a^1_4}\} \big), \big(\{e_2\}, \{\frac{0.90}{a^1_2}, \frac{0.90}{a^1_3}, \frac{0.10}{a^1_4}\} \big) \Big\};$$

$$CS^1_2 = \Big\{ \big(\{e_4, e_5, e_6, e_7\}, \{\frac{0.18}{a^1_2}, \frac{0.13}{a^1_3}, \frac{0.83}{a^1_4}\} \big) \Big\};$$

$$CS^2_1 = \Big\{ \big(\{e_1, e_2, e_3\}, \{\frac{0.86}{a^2_2}, \frac{0.80}{a^2_3}\} \big) \Big\};$$

$$CS^2_2 = \Big\{ \big(\{e_4, e_5, e_6, e_7\}, \{\frac{0.15}{a^2_2}, \frac{0.12}{a^2_3}\} \big) \Big\};$$

$$CS^3_1 = \Big\{ \big(\{e_1, e_2, e_3\}, \{\frac{0.84}{a^3_2}\} \big) \Big\};$$

$$CS^3_2 = \Big\{ \big(\{e_4, e_5, e_6, e_7\}, \{\frac{0.12}{a^3_2}\} \big) \Big\}.$$

3.4. Scale collaboration and uncertainty assessment

This subsection explores the collaborative mechanisms of progressive concepts at various scales and completes classification task using concepts as knowledge carriers. First, the following discussion is based on the two assumptions:

(1) Progressive concepts can reflect the essential characteristics of data;
(2) The representational power of progressive concepts is stronger than that of data.

According to the discussion in Section 3.3, for any $l \in \{1, 2, \dots, L\}$, we can learn the concept set family $CS_l = \{CS^l_k | k = 1, 2, \dots, K\}$ from F'_l. It can be observed that the progressive concepts can be seen as a more powerful method of knowledge representation that is abstracted from data. The intents of progressive concepts along with the corresponding labels under the lth scale are denoted as Int_l for simplicity, and define $Int = \{Int_1, Int_2, \dots, Int_L\}$. An example is provided below for further clarification.

Example 4. Continued from Example 3. The intents of all progressive concepts learned from F'_1, F'_2, and F'_3, along with their labels, are listed

Table 5
The intents of progressive concepts and their labels under the first scale.

Extent	a_2^1	a_3^1	a_4^1	d_1	d_2
e_1, e_3	0.78	0.10	0.18	1	0
e_2	0.90	0.90	0.10	1	0
e_4, e_5, e_6, e_7	0.18	0.13	0.83	0	1

Table 6
The intents of progressive concepts and their labels under the second scale.

Extent	a_2^2	a_3^2	d_1	d_2
e_1, e_2, e_3	0.86	0.80	1	0
e_4, e_5, e_6, e_7	0.15	0.12	0	1

Table 7
The intents of progressive concepts and their labels under the third scale.

Extent	a_2^3	d_1	d_2
e_1, e_2, e_3	0.84	1	0
e_4, e_5, e_6, e_7	0.12	0	1

in Tables 5–7, i.e., Int_1, Int_2 and Int_3 correspond to Tables 5–7, respectively.

For convenience, the feature part and the label part of dataset $F'_l \in MF'$ are denoted as $Fea(F'_l)$ and $Lab(F'_l)$, respectively. In subsequent discussions, we aim to give the labels of new objects. To achieve this, we need to introduce a classifier, denoted as CL, which could be K-Nearest Neighbors, Decision Tree or else. It should be noted that the classifiers at all scales should be the same type. As is well known, the train feature, train label and the test feature are the necessary inputs for a classifier, and the output is a predicted category Y. It can be described as

$$Y \leftarrow CL(\{\text{train feature,train label}\}|\text{test feature}). \tag{21}$$

We provide a detailed description of the detailed process. The multi-scale dataset after feature selection is denoted by $MF' = \{F'_1, F'_2, \dots, F'_L\}$. The intents of progressive concept with corresponding labels are represented by $Int = \{Int_1, Int_2, \dots, Int_L\}$. A new object o is characterized by its features $T_o = \{I(o, a_n^l)|a_n^l \in A'_l, l = 1, 2, \dots, L\}$. Then, the label of o is given based on CL, MF' and Int. To this end, two questions are considered here: (1) How to collaborate on integrating the progressive concepts across various scales to predict the label of object o; (2) How to address the representational disparities between progressive concepts and data.

For question (1), the label of object o is determined as follows:

$$Y \leftarrow \sum_{l=1}^{L} W_l \cdot CL(Int_l|T_o). \tag{22}$$

For question (2), we assess the uncertainty of data within each scale by evaluating the differences between progressive concepts and data, and generate a weight vector W as follows:

$$W = \{W_1, W_2, \dots, W_L\}. \tag{23}$$

The weight W is calculated using the following equation:

$$W_l = \exp\left(\frac{2 \times count_l}{\sum_{l=1}^{L} count_l}\right). \tag{24}$$

Here, $count_l$ refers to the number of correct predictions obtained by comparing the results from the classifier CL, which is trained on Int_l and tested on $Fea(F'_l)$, against the ground truth $Lab(F'_l)$, i.e.,

$$count_l \leftarrow Lab(F'_l) \leftarrow CL(Int'_l|Fea(F'_l)). \tag{25}$$

Some explanations are provided here for the convenience of the readers' understanding.

(1) In Eq. (22), $CL(Int_l|T_o)$ represents the label in vector form. For example, suppose $D = \{d_1, d_2, d_3\}$, if the predicted label is $d_1 \in D$, then it is represented as $(1, 0, 0)$. Suppose there are two scales and $W = \{1.5, 1.2\}$, where 1.5 and 1.2 represents the weight of the data at the first and the second scale, respectively. For an object o, if it is predicted that the labels of o is d_1 and d_2 at the first and second scales respectively, then it can be calculated that $1.5 \times (1, 0, 0) + 1.2 \times (0, 1, 0) = (1.5, 1.2, 0)$. The label corresponding to the maximum value in $(1.5, 1.2, 0)$ is the label of o, indicating that the label of o is d_1.
(2) The weights of each scale are calculated based on the differences in the representational capabilities between intents of the progressive concepts and the data. After learning the progressive concepts from the data, we take the intents of the progressive concepts as the training set, while the data itself serves as the testing set under each scale. We then compare the prediction results with the true labels of the data to assess the uncertainty of the data at each scale, as represented by the weight vector W.

The content discussed in this subsection corresponds to Algorithm 4. Overall, the proposed model is divided into three stages: generating the multi-scale dataset, advancing feature selection and learning progressive concepts, and collaboratively predicting the labels for new objects based on progressive concepts across multiple scales. The time complexity of the first stage is $O(L|A||E|)$, and the second stage is $O(L^2|A|^2|E| + L|A||E|^2)$. Therefore, the total time complexity of learning progressive concepts from a multi-scale dataset is $O(L^2|A|^2|E| + L|A||E|^2)$.

Algorithm 4: Scale collaboration mechanism for object classification.

Input: A concept space CS, a multi-scale dataset $MF = \{F'_1, F'_2, \dots, F'_L\}$, a new object o with feature $\{I(o, a_n^l)|a_n^l \in A'_l, l = 1, 2, \dots, L\}$, a classifier CL.
Output: The label of o.
for $l \in \{1, 2, \dots, L\}$ **do**
  Calculate Int_l based on CS_l;
  Calculate $count_l$ based on Eq. (25);
Calculate the weight vector W based on Eq. (24);
Calculate Y based on Eq. (22);
Assign the label corresponding to the maximum value of Y to o.

3.5. Dynamic knowledge updating via concepts

One of the key characteristics of CCL is its ability to flexibly handle dynamic data. This subsection explores the dynamic updating mechanisms of knowledge based on the proposed collaborative multi-scale concept-cognitive learning model.

We formally describe the problem of dynamic knowledge updating. The knowledge in this paper is composed of three parts: the multi-scale dataset MF', the progressive concepts Int at multiple scales, and the uncertainty measurement vector W. Then, when a new set of objects, denoted as Chunk, is added, the mechanism of knowledge updating is studied. For any object $o \in$ Chunk, its feature values $\{I(o, a_n^l)|a_n^l \in A'_l, l = 1, 2, \dots, L\}$ are known, while its label is unknown. More specifically, based on Chunk, we need to obtain the updated multi-scale dataset $\Delta MF'$, the updated progressive concepts ΔInt and the updated weight ΔW. The detailed steps are shown in Algorithm 5.

4. Experiments

To validate the effectiveness of the proposed collaborative multi-scale concept-cognitive learning model (CM-CCL), several experiments are conducted, and the results are discussed in this section. The experiments are designed to draw the following three conclusions: (1) Concepts exhibit superior representation capabilities compared to raw data;

Algorithm 5: Knowledge updating in dynamic environment.

Input: The initial multi-scale dataset MF', the initial intents of progressive concepts Int, the updated set of object Chunk and the weight vector W.
Output: The updated multi-scale dataset $\Delta MF'$, the updated intents of progressive concepts ΔInt, and the updated weight vector ΔW.
Assign labels to all objects in Chunk based on Algorithm 4;
Add the feature values and labels of each object to MF' to obtain $\Delta MF'$;
Learn the progressive concepts based on Algorithm 3 from $\Delta MF'$ and get the updated ΔInt;
Based on ΔInt, $\Delta MF'$, and Eqs. (24) and (25), obtain the updated ΔW;
return $\Delta MF'$, ΔInt and ΔW.

Table 8
Detailed information of the datasets.

ID	Dataset	Object	Feature	Class
1	Breast Cancer Coimbra	116	9	2
2	Darwin	174	450	2
3	Glass Identification	214	9	6
4	Heart Failure Clinical Records	299	12	2
5	Haberman	306	3	2
6	Indian Liver Patient	579	9	2
7	Australia	690	14	2
8	Parkinson's Disease Classification	756	753	2
9	HCV	1385	28	4
10	Steel Plates Faults	1941	27	7
11	CTG	2126	30	3
12	Dry Bean	13,611	16	7

(2) The CM-CCL model outperforms the single-scale CCL model; (3) The CM-CCL model demonstrates better performance in dynamic environment.

4.1. Experiment settings

The experiments are conducted on 12 datasets from the UCI repository, and the detailed information about these datasets are listed in Table 8.

The datasets used in this paper are all continuous data. In order to construct fuzzy data and to meet the requirements of the comparison methods used in this paper, we scale the values of all features to the range of $[0.1000, 0.9000]$ based on the following equation:

$$\widetilde{I}(e,a) = 0.1 + (0.9-0.1)\frac{I'(e,a)-\min(a)}{\max(a)-\min(a)}, \tag{26}$$

where e is an object, and $\min(a)$ and $\max(a)$ are the minimum and maximum of a. $I'(e,a)$ represents the value of object e under feature a in the original dataset.

The experiments are conducted on a personal computer with OS: Microsoft WIN10; Processor: Intel(R) Core(TM) i7-13700H CPU @2400 Mhz; Memory: 32 GB; Programming language: MATLAB R2020b.

4.2. Baseline comparison

This subsection aims to demonstrate that the representational power of concepts derived from multi-scale datasets is stronger than that of raw data. To illustrate this, we conducted experiments using three representative classifiers: K-Nearest Neighbors (KNN), Decision Trees(DT), and Random Forests(RF). On one hand, we input the data directly into the classifiers to obtain the corresponding classification results. On the other hand, we embedded the classifiers into our CM-CCL framework to achieve classification accuracy. To ensure fairness, the experiments in this subsection utilize a ten-fold cross-validation approach with consistent data partitioning, and the average classification accuracy and standard deviation are recorded in percentage form. The results are presented in Table 9. The CM-CCL methods under KNN, DT, and RF classifiers are denoted as CM-CCL-KNN, CM-CCL-DT, and CM-CCL-RF, respectively. Additionally, the classification performance using the raw data directly with the KNN, DT, and RF classifiers is recorded for comparison. Furthermore, the average classification accuracies that show superior performance across each dataset are highlighted in bold for each classifier. Under each dataset, the parameter δ in the proposed multi-scale concept cognitive learning model is searched with a step size of 0.1 from the interval [0, 1]. Besides, the number of scales in CM-CCL is set to 5 for all datasets.

According to Table 9, it is evident that the performance of CM-CCL-KNN surpasses that of KNN in 10 datasets, CM-CCL-DT exceeds DT in 11 datasets, and CM-CCL-RF outperforms RF in 10 datasets, indicating that the progressive concepts have better representational capability compared to the data. The experimental results demonstrate that the representational power of progressive concepts learned across multiple scales is stronger than that of the data, thereby validating conclusion (1).

4.3. Comparison with other methods

In this subsection, some comparative experiments are conducted to verify the effectiveness of CM-CCL. Four CCL models and a representation learning method are adopted as comparative methods. Four CCL models are, respectively, incremental learning mechanism based on progressive fuzzy three-way concept (ILMPFTC) [10], dynamic updating mechanism of progressive weighted fuzzy concept (DMPWFC) [12], memory-based fuzzy concept-cognitive learning (MFCCL) [15], and interval-intent fuzzy concept re-cognition learning model(IFCRL) [13]. An association-based fusion method (AF) [33] is also employed. AF is a represent learning method, and the same classifiers, KNN, DT, and RF, are used as the base classifiers, denoted as AF-KNN, AF-DT and AF-RF, respectively. All parameter configurations are set according to the original literature. Similarly, under the 12 datasets, after ten-fold cross-validation, the average classification accuracy and standard deviation for each algorithm is recorded in Table 10. And the average rank of each method is listed in the last row of Table 10.

From Table 10, some experimental results can be observed. Firstly, CM-CCL-RF achieve the best classification performance in 8 out of 12 datasets, and CM-CCL-KNN achieve the best classification performance in 1 datasets. Secondly, compared to AF, CM-CCL-RF outperforms AF-RF on 9 datasets, CM-CCL-DT outperforms AF-DT on 10 datasets, and CM-CCL-KNN outperforms AF-KNN on 9 datasets. Furthermore, CM-CCL outperform single-scale CCL methods across most datasets. More specifically, the average rank of CM-CCL-RF, CM-CCL-DT, CM-CCL-KNN are, respectively, 2.33, 4.50 and 4.79, which are all better than those of ILMPFTC (6.00), DMPWFC (8.38), MFCCL (4.83) and IFCRL (9.25).

The Friedman test at a significance level of 0.05 is conducted to further illustrate the advantages of CM-CCL [34]. The calculated p-value is 8.2937×10^{-9}, which is less than 0.05, indicating that there are significant differences between the different methods at the 0.05 significance level. Furthermore, the Nemenyi post hoc test is conducted to further analyze the differences between pairs of methods [35]. The critical distance is calculated by

$$\mathrm{CD} = q_{0.05}\sqrt{\frac{k(k+1)}{6N}}, \tag{27}$$

where $q_{0.05} = 2.773$, $k = 10$ and $N = 12$. The CD value is calculated to be 3.4275. Subsequently, the corresponding CD plot is displayed in Fig. 2. It can be observed that the performance of CM-CCL-RF is superior to that of CM-CCL-DT, which in turn outperforms CM-CCL-KNN. Moreover, the average rank of CM-CCL-KNN is still better than that of the four single-scale CCL models.

Table 9
Classification accuracy comparison between data and concepts.

Dataset	CM-CCL-KNN	KNN	CM-CCL-DT	DT	CM-CCL-RF	RF
1	**68.48** ± 18.06	64.47 ± 12.01	69.09 ± 12.55	**74.39** ± 14.10	**74.17** ± 10.14	69.96 ± 15.59
2	**78.76** ± 8.14	71.93 ± 13.21	**85.56** ± 10.80	76.27 ± 10.30	**90.13** ± 9.28	86.63 ± 8.86
3	**69.65** ± 7.20	68.79 ± 10.34	**66.93** ± 13.38	66.00 ± 13.07	**76.75** ± 9.81	76.73 ± 6.94
4	70.25 ± 3.85	**71.26** ± 6.79	**80.97** ± 6.99	77.93 ± 7.04	**83.62** ± 7.09	83.28 ± 7.86
5	**73.83** ± 8.13	68.99 ± 9.22	**73.52** ± 9.23	69.39 ± 10.41	**68.24** ± 8.29	66.43 ± 9.60
6	**71.50** ± 4.95	67.88 ± 5.83	**70.81** ± 6.04	65.63 ± 7.83	**72.88** ± 4.66	70.64 ± 5.50
7	**84.35** ± 3.33	83.19 ± 2.66	**83.04** ± 3.28	82.61 ± 3.86	85.07 ± 4.16	**86.96** ± 4.32
8	**91.40** ± 3.67	91.13 ± 2.95	**82.02** ± 5.10	81.88 ± 4.04	**88.09** ± 2.93	87.69 ± 2.94
9	**25.49** ± 2.86	23.61 ± 4.15	**25.78** ± 4.43	23.33 ± 4.20	**26.35** ± 3.06	25.42 ± 4.33
10	68.57 ± 4.06	**70.32** ± 4.63	**74.91** ± 3.22	73.37 ± 3.47	**79.29** ± 2.35	77.90 ± 2.33
11	**90.97** ± 1.96	90.26 ± 2.18	**93.18** ± 1.67	93.13 ± 1.29	**95.39** ± 1.50	95.25 ± 1.70
12	**91.65** ± 0.57	**91.65** ± 0.66	**90.43** ± 0.96	89.79 ± 1.02	92.48 ± 0.81	**92.56** ± 0.94

Table 10
Classification accuracy comparison with other methods.

Dataset	CM-CCL-RF	AF-RF	CM-CCL-DT	AF-DT	CM-CCL-KNN	AF-KNN	ILMPFTC	DMPWFC	MFCCL	IFCRL
1	74.17 ± 10.14	77.65 ± 13.71	69.09 ± 12.55	**78.41** ± 13.49	68.48 ± 18.06	69.34 ± 15.96	73.18 ± 12.77	57.80 ± 11.43	75.83 ± 11.72	49.74 ± 15.91
2	**90.13** ± 9.28	86.18 ± 9.59	85.56 ± 10.8	83.76 ± 9.09	78.76 ± 8.14	81.62 ± 5.07	73.07 ± 11.74	33.40 ± 8.42	78.79 ± 7.57	50.03 ± 8.52
3	**76.75** ± 9.81	70.13 ± 7.26	66.93 ± 13.38	64.57 ± 0.28	69.65 ± 7.20	69.72 ± 7.59	68.40 ± 12.76	50.61 ± 13.50	68.40 ± 12.76	30.77 ± 17.76
4	**83.62** ± 7.09	78.95 ± 6.98	80.97 ± 6.99	73.97 ± 8.21	70.25 ± 3.85	69.59 ± 8.31	65.57 ± 8.38	69.94 ± 6.54	70.92 ± 6.22	49.92 ± 19.44
5	68.24 ± 8.29	70.66 ± 8.31	73.52 ± 9.23	70.32 ± 7.62	**73.83** ± 8.13	69.63 ± 8.58	72.57 ± 12.03	39.88 ± 7.00	73.20 ± 10.65	69.97 ± 16.02
6	**72.88** ± 4.66	72.19 ± 6.30	70.81 ± 6.04	69.94 ± 5.38	71.50 ± 4.95	69.78 ± 4.31	64.94 ± 6.60	62.19 ± 6.91	64.94 ± 6.60	62.54 ± 19.07
7	**85.07** ± 4.16	84.78 ± 3.68	83.04 ± 3.28	81.16 ± 3.58	84.35 ± 3.33	83.77 ± 2.44	79.42 ± 4.03	84.35 ± 3.91	79.42 ± 4.03	48.12 ± 8.45
8	88.09 ± 2.93	86.89 ± 3.12	82.02 ± 5.10	81.46 ± 5.00	91.40 ± 3.67	82.81 ± 4.66	**95.37** ± 2.10	35.17 ± 7.49	**95.37** ± 2.10	64.07 ± 21.53
9	26.35 ± 3.06	26.42 ± 1.48	25.78 ± 4.43	**26.49** ± 5.12	25.49 ± 2.86	25.34 ± 2.95	24.95 ± 4.41	25.85 ± 3.57	26.14 ± 4.30	25.71 ± 3.21
10	**79.29** ± 2.35	73.26 ± 3.97	74.91 ± 3.22	66.20 ± 2.67	68.57 ± 4.06	66.82 ± 3.67	72.23 ± 3.34	32.20 ± 3.88	72.23 ± 3.34	24.27 ± 12.28
11	**95.39** ± 1.50	92.64 ± 1.66	93.18 ± 1.67	89.56 ± 2.48	90.97 ± 1.96	90.21 ± 2.13	91.63 ± 1.33	83.07 ± 1.86	91.44 ± 1.68	51.77 ± 33.77
12	**92.48** ± 0.94	90.62 ± 1.07	90.43 ± 0.96	87.72 ± 0.90	91.65 ± 0.57	88.83 ± 1.20	90.45 ± 0.50	49.74 ± 2.09	90.45 ± 0.50	15.27 ± 1.76
Average rank	**2.33**	2.83	4.50	5.67	4.79	6.42	6.00	8.38	4.83	9.25

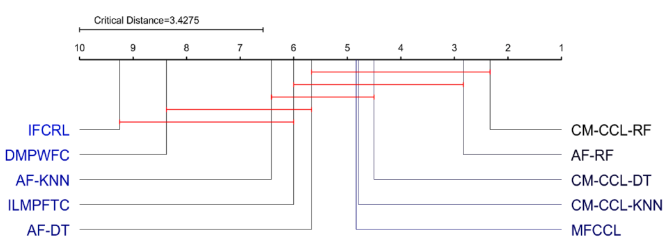

Fig. 2. Comparison between CM-CCL and other methods.

The experiments in this subsection confirm that the performance of CM-CCL is better than that of the single-scale CCL model, which validates conclusion (2).

4.4. Evaluation of dynamic concept learning performances

One of the major advantages of CCL is its ability to maintain good classification performance in dynamic environment. Therefore, in this subsection, we verify the performance of CM-CCL when objects are dynamically increases. In this paper, we assume a total of 5 incremental object sets. Specifically, we divide the raw data into 10 equal parts, selecting 5 of them as the initial training set to build the initial concept space. Each of the remaining parts is then used as an additional chunk incrementally. By comparing the labels generated by Algorithm 4 for each $Chunk_i$ with the true labels, we can obtain the classification accuracy for each $Chunk_i$.

It should be noted that ILMPFTC, DMPWFC, and MFCCL all provide methods for dynamic concept updating. We compare CM-CCL with these methods, and the corresponding experimental results are listed in Table 11. The classification accuracy for each chunk is recorded, along with the average value and standard deviation (Ave. + std.), and the rank of each algorithm for each dataset is calculated. The best average classification accuracy on each dataset is highlighted in bold.

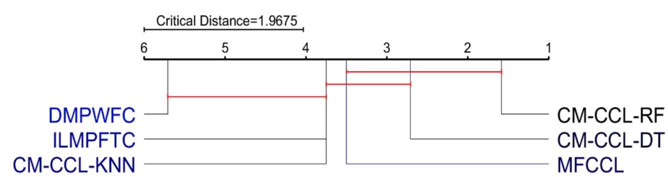

Fig. 3. Comparison between CM-CCl and single-scale CCL in dynamic environment.

Similarly, the Nemenyi post hoc test [35] is also conducted to verify the differences between each algorithm, which is presented in Fig. 3. From Fig. 3, it can be visually observed that the classification performance of CM-CCL-DT and CM-CCL-RF is superior to that of the four single-scale CCL methods. This indicates that in dynamic environments, CM-CCL can learn concepts with better representational capabilities compared to single-scale CCL methods, thus confirming conclusion (3).

4.5. Ablation experiment

Ablation experiments are conducted to verify the role of each step in CM-CCL in enhancing the representation capability of concepts. CM-CCL consists of three successive steps, namely the construction of multi-scale data, multi-scale feature selection, concept learning and integration for classification. As the first two steps determine the input feature quality for concept learning, we validated their necessity. The corresponding experimental results are shown in Table 12. Step 1 represents the construction of multi-scale data, and Step 2 indicates the multi-scale feature selection. ✓indicates the inclusion of the corresponding step, while × indicates its exclusion. The DT classifier is still used, and the average classification accuracy is recorded after ten-fold cross-validation. It is observed that Step 1 or Step 2 alone does not enhance the capability of concept representation. Only when both Step 1 and Step 2 are applied can the capability of concept representation be significantly improved.

Table 11
Classification accuracy comparison with CCL methods in dynamic environment.

Dataset	Methods	$Chunk_1$	$Chunk_2$	$Chunk_3$	$Chunk_4$	$Chunk_5$	Ave. ± std.	Rank	Dataset	Methods	$Chunk_1$	$Chunk_2$	$Chunk_3$	$Chunk_4$	$Chunk_5$	Ave. ± std.	Rank
1	CM-CCL-KNN	33.33	72.73	58.33	81.82	75.00	64.24 ± 19.28	5	7	CM-CCL-KNN	86.96	88.41	86.96	85.51	81.16	85.80 ± 2.79	2.5
	CM-CCL-DT	58.33	63.64	41.67	81.82	83.33	65.76 ± 17.37	4		CM-CCL-DT	79.71	84.06	79.71	85.51	76.81	81.16 ± 3.55	4
	CM-CCL-RF	58.33	72.73	75.00	81.82	66.67	70.91 ± 8.88	3		CM-CCL-RF	86.96	89.86	86.96	88.41	79.71	**86.38** ± 3.92	1
	ILMPFTC	75.00	54.55	83.33	72.73	83.33	73.79 ± 11.78	2		ILMPFTC	66.67	78.26	89.86	82.61	73.91	78.26 ± 8.76	5
	DMPWFC	41.67	54.55	83.33	63.64	66.67	61.97 ± 15.40	6		DMPWFC	88.41	81.16	92.75	82.61	84.06	85.80 ± 4.74	2.5
	MFCCL	58.33	81.82	83.33	72.73	83.33	**75.91** ± 10.77	1		MFCCL	66.67	78.26	89.86	82.61	72.46	77.97 ± 8.96	6
2	CM-CCL-KNN	88.24	72.22	66.67	83.33	88.24	79.74 ± 9.81	3	8	CM-CCL-KNN	88.00	88.16	76.00	82.89	81.33	83.28 ± 5.07	4
	CM-CCL-DT	76.47	83.33	94.44	83.33	88.24	85.16 ± 6.67	2		CM-CCL-DT	74.67	80.26	77.33	78.95	82.67	78.78 ± 3.01	5
	CM-CCL-RF	88.24	83.33	94.44	100.00	100.00	**93.20** ± 7.35	1		CM-CCL-RF	81.33	84.21	82.67	82.89	85.33	83.29 ± 1.53	3
	ILMPFTC	64.71	61.11	77.78	66.67	76.47	69.35 ± 7.39	5		ILMPFTC	90.67	93.42	85.33	88.16	85.33	**88.58** ± 3.50	1.5
	DMPWFC	23.53	22.22	50.00	27.78	23.53	29.41 ± 11.70	6		DMPWFC	44.00	36.84	46.67	46.05	38.67	42.45 ± 4.44	6
	MFCCL	82.35	66.67	77.78	88.89	82.35	79.61 ± 8.25	4		MFCCL	90.67	93.42	85.33	88.16	85.33	**88.58** ± 3.50	1.5
3	CM-CCL-KNN	52.38	63.64	66.67	71.43	63.64	63.55 ± 7.01	3	9	CM-CCL-KNN	22.46	23.19	26.81	20.14	31.16	24.75 ± 4.31	5
	CM-CCL-DT	71.43	40.91	61.90	80.95	63.64	63.77 ± 14.83	2		CM-CCL-DT	26.81	20.29	27.54	29.50	26.09	26.04 ± 3.46	3
	CM-CCL-RF	61.90	45.45	7143	90.48	72.73	**68.40** ± 16.46	1		CM-CCL-RF	29.71	32.61	34.06	23.74	18.12	27.65 ± 6.63	2
	ILMPFTC	57.14	45.45	61.90	71.43	68.18	60.82 ± 10.22	4.5		ILMPFTC	20.29	19.57	31.88	25.90	31.88	25.90 ± 5.98	4
	DMPWFC	42.86	31.82	71.43	57.14	36.36	47.92 ± 16.25	6		DMPWFC	20.29	28.26	27.54	23.74	21.74	24.31 ± 3.50	6
	MFCCL	57.14	45.45	61.92	71.43	68.18	60.82 ± 10.22	4.5		MFCCL	24.64	28.26	30.43	24.46	31.88	**27.94** ± 3.35	1
4	CM-CCL-KNN	63.33	63.33	70.00	73.33	63.33	66.67 ± 4.71	5	10	CM-CCL-KNN	67.01	72.68	69.07	67.53	65.98	68.45 ± 2.61	5
	CM-CCL-DT	80.00	76.67	76.67	86.67	83.33	80.67 ± 4.35	2		CM-CCL-DT	72.16	76.80	66.49	75.26	70.62	72.27 ± 4.05	2
	CM-CCL-RF	86.67	80.00	80.00	96.67	93.33	**87.33** ± 7.60	1		CM-CCL-RF	79.38	83.51	77.84	78.87	69.07	**77.73** ± 5.30	1
	ILMPFTC	66.67	73.33	66.67	56.67	76.67	68.00 ± 7.67	4		ILMPFTC	67.01	73.20	69.59	72.16	67.53	69.90 ± 2.74	3.5
	DMPWFC	56.67	73.33	66.67	66.67	63.33	65.33 ± 6.06	6		DMPWFC	40.21	37.63	38.14	41.75	34.02	38.35 ± 2.93	6
	MFCCL	63.33	63.33	70.00	83.33	80.00	72.00 ± 9.31	3		MFCCL	67.01	73.20	69.59	72.16	67.53	69.90 ± 2.74	3.5
5	CM-CCL-KNN	66.67	67.74	74.19	83.33	87.10	**75.81** ± 9.16	1.5	11	CM-CCL-KNN	85.45	88.21	88.26	87.26	90.14	87.86 ± 1.71	5
	CM-CCL-DT	66.67	67.74	74.19	83.33	87.10	**75.81** ± 9.16	1.5		CM-CCL-DT	92.49	94.81	92.02	95.28	90.14	92.95 ± 2.11	2
	CM-CCL-RF	63.33	67.74	74.19	73.33	83.87	72.49 ± 7.74	3		CM-CCL-RF	93.90	94.34	93.90	97.17	94.37	**94.73** ± 1.38	1
	ILMPFTC	43.33	70.97	64.52	70.00	80.65	65.89 ± 13.88	5		ILMPFTC	88.73	91.51	91.08	91.98	92.02	91.06 ± 1.36	3
	DMPWFC	40.00	35.48	41.94	40.00	32.26	37.96 ± 3.96	6		DMPWFC	78.40	83.49	82.16	84.43	84.98	82.69 ± 2.63	6
	MFCCL	50.00	61.29	61.29	70.00	87.10	65.94 ± 13.80	4		MFCCL	88.26	91.51	91.08	91.98	92.02	90.97 ± 1.56	4
6	CM-CCL-KNN	75.44	67.24	58.62	62.07	68.97	66.47 ± 6.49	4	12	CM-CCL-KNN	91.55	91.40	91.70	91.26	90.60	91.30 ± 0.43	2
	CM-CCL-DT	63.16	79.31	68.97	74.14	58.62	68.84 ± 8.28	2		CM-CCL-DT	90.96	90.52	91.04	90.30	88.98	90.36 ± 0.83	3
	CM-CCL-RF	73.68	79.31	68.97	74.14	68.97	**73.01** ± 4.31	1		CM-CCL-RF	92.14	92.58	92.73	92.58	90.67	**92.14** ± 0.85	1
	ILMPFTC	71.93	72.41	67.24	63.79	58.62	66.80 ± 5.79	3		ILMPFTC	90.01	91.04	90.08	89.71	88.76	89.92 ± 0.82	4.5
	DMPWFC	54.39	58.62	55.17	55.17	74.14	59.50 ± 8.35	6		DMPWFC	49.89	50.77	51.95	49.74	50.99	50.67 ± 0.90	6
	MFCCL	28.07	79.31	70.69	70.69	70.69	63.89 ± 20.37	5		MFCCL	90.01	91.04	90.08	89.71	88.76	89.92 ± 0.82	4.5

Table 12
Ablation study in the steps in CM-CCL.

Step 1	Step 2	Dataset 1	Dataset 2	Dataset 3	Dataset 4	Dataset 5	Dataset 6	Dataset 7	Dataset 8	Dataset 9	Dataset 10	Dataset 11	Dataset 12
×	×	64.39±17.38	76.27±10.30	**67.38±13.54**	77.95±7.00	36.00±8.31	69.94±5.50	81.01±4.66	81.87±4.23	23.40±4.25	72.28±2.87	93.23±1.73	89.91±0.81
✓	×	62.58±20.97	76.86±10.47	67.38±13.91	77.97±8.14	36.98±7.99	67.53±7.19	82.46±4.24	81.74±4.27	23.26±4.18	72.39±3.31	**93.37±1.77**	89.90±0.81
✓	✓	**69.09±12.55**	**85.56±10.80**	66.93±13.38	**80.97±6.99**	**73.52±9.23**	**70.81±6.04**	**83.04±3.28**	**82.02±5.10**	**25.78±4.43**	**74.91±3.22**	93.18±1.67	**90.43±0.96**

Moreover, the proposed feature selection method in this paper can also be replaced. To illustrate its effectiveness, we also tried another filter feature selection method based on the Pearson correlation coefficient (PCC) [36], and the corresponding code is available.[1] It should be pointed out that PCC first sorts the features and then specifies the number of features to be selected. For the sake of fairness, under each dataset, we successively select 50 %, 55 %, …, 100 % of the features and take the best classification performance among them. The experimental results are recorded in Table 13, where "Baseline" indicates the classification performance without feature selection under the CM-CCL framework. It can be seen that the feature selection method designed in this paper achieves the best results on 9 datasets. In addition, PCC outperforms the Baseline on 11 datasets, which indicates that the proposed CM-CCL framework is scalable and that the feature selection method can also be replaced by other methods to enhance the representation capability of learned concepts.

[1] https://github.com/JingweiToo/Filter-Feature-Selection-Toolbox

4.6. Parameter analysis

A parameter δ is employed in CM-CCL. To analysis the impact of δ, some experiments are conducted. Under the classifier DT, we adjusted δ within the range [0,1] with a step size of 0.1. After ten-fold cross-validation, the average classification accuracy and standard deviation for different parameter settings are recorded in Table 14, and the corresponding data are intuitively presented in Fig. 4. The number of selected features at each scale is determined by δ, and as δ gradually increases, the number of selected features at each scale will also increase step by step. When $\delta = 1$, all feature will be selected. Overall, the optimal δ value for achieving the best classification accuracy in each dataset is less than 1, which indicates that the multi-scale feature selection method is effective in enhancing the representation capability of concepts.

It should be pointed out that the characteristics of using concepts to represent knowledge in data are reflected in two aspects: (1) Precision: Concepts capture knowledge more accurately than raw data, as evidenced by classification accuracy.(2) Conciseness: A smaller number of concepts can be used to represent the valuable information in data,

Table 13
Ablation study on the feature selection method.

Method	Dataset 1	Dataset 2	Dataset 3	Dataset 4	Dataset 5	Dataset 6	Dataset 7	Dataset 8	Dataset 9	Dataset 10	Dataset 11	Dataset 12
Baseline	62.58±20.97	76.86±10.47	67.38±13.91	77.97±8.14	36.98±7.99	67.53±7.19	82.46±4.24	81.74±4.27	23.26±4.18	72.39±3.31	93.37±1.77	89.90±0.81
PCC	67.27±9.32	76.90±10.70	**72.03±13.93**	80.94±6.47	**73.52±9.23**	68.39±7.26	82.46±4.24	**83.20±3.79**	24.19±2.68	72.70±3.29	**94.03±1.46**	90.03±0.87
Our	**69.09±12.55**	**85.56±10.80**	66.93±13.38	**80.97±6.99**	**73.52±9.23**	**70.81±6.04**	**83.04±3.28**	82.02±5.10	**25.78±4.43**	**74.91±3.22**	93.18±1.67	**90.43±0.96**

Table 14
Classification accuracy under the classifier DT for different values of δ.

Dataset	$\delta = 0$	$\delta = 0.1$	$\delta = 0.2$	$\delta = 0.3$	$\delta = 0.4$	$\delta = 0.5$	$\delta = 0.6$	$\delta = 0.7$	$\delta = 0.8$	$\delta = 0.9$	$\delta = 1$
1	54.55±14.40	54.55±14.40	54.55±14.40	55.38±14.88	55.38±14.88	55.38±14.88	66.52±18.60	**69.09±12.55**	64.55±17.16	66.36±13.16	60.91±19.76
2	68.99±9.08	**85.56±10.80**	83.20±8.57	77.35±11.10	78.46±15.94	77.91±16.27	77.91±16.27	77.42±13.38	75.75±11.99	76.34±11.52	76.93±11.34
3	32.71±11.61	32.71±11.61	34.13±10.58	47.84±13.36	54.33±9.30	60.41±10.45	60.41±10.45	60.89±9.90	66.47±14.72	**66.93±13.38**	65.02±13.27
4	67.92±4.61	80.30±7.72	80.30±7.72	80.30±7.72	80.30±7.72	80.30±7.72	**80.97±6.99**	79.62±6.70	79.62±6.88	79.62±7.40	77.63±8.41
5	**73.52±9.23**	73.52±9.23	73.52±9.23	73.52±9.23	73.52±9.23	73.52±9.23	71.26±14.05	71.26±14.05	70.94±14.89	36.98±7.99	36.98±7.99
6	39.55±12.53	40.76±10.71	42.66±10.56	42.66±10.56	63.22±15.82	63.22±15.82	68.40±6.01	69.78±5.52	**70.81±6.04**	69.78±6.84	67.36±7.29
7	51.74±8.48	57.10±16.58	57.10±16.58	66.67±12.98	79.71±4.37	81.59±3.49	82.03±2.66	82.03±2.66	**83.04±3.28**	83.04±3.42	82.17±3.99
8	57.20±15.65	**82.02±5.10**	81.21±4.79	81.21±4.79	81.21±4.79	81.21±4.79	81.21±4.79	81.21±4.79	81.21±4.79	81.21±4.79	81.21±4.79
9	24.55±2.93	25.06±4.05	25.06±4.05	25.42±4.24	**25.78±4.43**	25.78±4.43	25.78±4.43	25.78±4.43	23.76±4.54	23.76±4.54	23.47±4.40
10	7.63±3.32	72.49±2.83	72.95±2.82	74.45±3.03	74.65±3.70	74.50±3.78	**74.91±3.22**	74.86±3.28	74.75±3.11	74.50±3.58	72.80±3.39
11	57.20±12.96	89.27±3.50	**93.18±1.67**	93.18±1.67	93.18±1.67	93.18±1.67	93.18±1.67	93.18±1.67	93.18±1.67	93.18±1.67	93.18±1.67
12	10.00±3.30	73.59±6.80	81.81±1.07	81.81±1.07	82.24±0.78	82.24±0.78	84.23±1.15	89.30±0.83	90.35±0.80	**90.43±0.96**	90.09±0.90

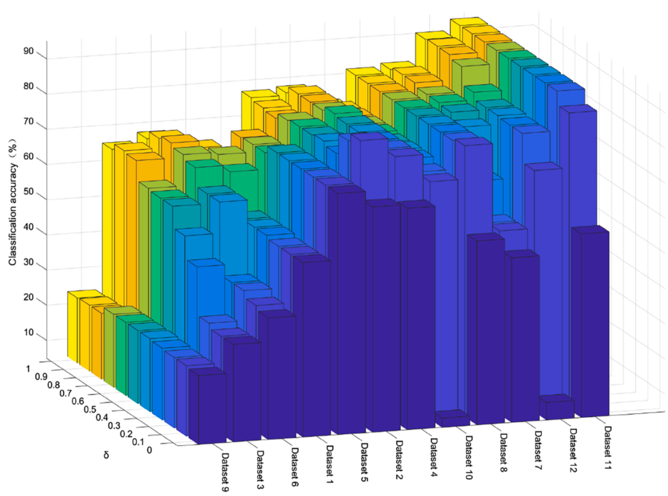

Fig. 4. Variation of classification accuracy with parameter δ across 12 datasets.

that is, the number of concepts is less than the number of objects. To verify (2), under the lth scale, we define the compression rate (CR_l) as follows:

$$CR_l = 1 - \frac{\sum_{k=1}^{K} |CS_k^l|}{|E|}. \tag{28}$$

The larger the CR_l, the fewer the number of concepts learned at the lth scale compared to the number of objects, which means that the learned concepts have stronger representational power. Under the optimal parameter δ for each dataset, we verified the CR_l value for each scale, presented in the form of percentages. The results are listed in Table 15, and the last column indicates the running time of CM-CCL in seconds. It can be seen that the CR_l values on the 10 datasets are greater than 0, indicating that the number of concepts learned on these datasets is less than the number of objects. Thus, CM-CCL represents knowledge in data using fewer concepts than the number of objects, achieving better classification performance. Besides, as the scale gradually increases, the compression rate also gradually rises, which means that more important information is being presented at bigger scales.

In addition, the experiments specified the number of scales as 5. This part further analyzes the classification performance as the number of scales varies from 1 to 5, and DT is still used as the basic classifier.

Table 15
Compression rate at each scale.

Dataset	CR_1	CR_2	CR_3	CR_4	CR_5	Time
1	22.86	58.10	86.67	95.24	95.24	0.89
2	0.00	0.64	3.18	6.37	21.66	2.19
3	1.04	1.04	1.04	1.55	4.66	1.55
4	18.59	57.25	57.25	72.49	72.49	1.24
5	94.93	98.55	98.55	99.28	99.28	1.84
6	62.76	62.76	69.10	77.93	88.48	2.21
7	65.38	65.38	67.63	76.17	76.17	2.27
8	0.00	0.00	0.00	0.00	0.00	80.05
9	0.00	0.00	0.00	0.00	0.00	8.36
10	1.37	1.37	1.43	1.43	2.92	10.76
11	4.75	4.75	4.75	5.22	16.09	17.71
12	0.45	0.45	0.45	55.15	55.15	229.08

Table 16
Variation of classification accuracy with parameter L under DT classifier.

Dataset	$L = 1$	$L = 2$	$L = 3$	$L = 4$	$L = 5$
1	68.26±13.93	68.26 ± 13.93	63.48 ± 20.78	**71.59 ± 11.24**	69.09± 12.55
2	76.37±8.51	82.65 ± 6.35	84.38 ± 8.89	84.93 ± 9.36	**85.56± 10.80**
3	**72.06±13.37**	71.58 ± 14.03	65.61 ± 14.53	66.45 ± 13.29	66.93± 13.38
4	78.29±7.01	79.29 ± 6.38	78.29 ± 7.19	78.62 ± 6.64	**80.97± 6.99**
5	47.63±20.27	**73.52 ± 9.23**	73.52 ± 9.23	73.52 ± 9.23	73.52± 9.23
6	69.43±5.66	69.43 ± 5.66	**71.50 ± 6.05**	71.50 ± 6.05	70.81± 6.04
7	80.72±4.53	82.61 ± 4.16	83.04 ± 4.04	**83.48 ± 3.22**	83.04± 3.28
8	81.60±3.92	81.47 ± 3.45	81.34 ± 4.32	81.34 ± 4.65	**82.02± 5.10**
9	25.49±3.90	23.69 ± 4.45	25.35 ± 4.04	25.50 ± 4.08	**25.78± 4.43**
10	72.95±2.91	73.42 ±2.68	74.60 ± 3.75	74.70 ± 3.60	**74.91± 3.22**
11	93.13±1.60	93.08±1.56	93.23 ±1.73	**93.27 ± 1.71**	93.18± 1.67
12	89.93±0.74	89.96 ±0.73	90.06 ±0.85	**90.59 ± 0.86**	90.43± 0.96

The experimental results are presented in Table 16. As evidenced by the experimental results, in most cases, higher classification accuracy is achieved when there are more scales. This demonstrates that the proposed CM-CCL model is capable of fully mining the multi-scale information of features to enhance the representation ability of concepts.

4.7. Analysis of the effects of embedding different CCL models in the CM-CCL framework

The proposed CM-CCL is a multi-scale concept-cognitive learning framework that can incorporate existing CCL models. To verify the effectiveness and scalability of CM-CCL, in this subsection, MFCCL and ILMPFTC are embedded in CM-CCL. Specifically, the concept learning

Table 17
Classification accuracy of embedding ILMPFTC in CM-CCL.

Dataset	CM-CCL-KNN	CM-CCL-DT	CM-CCL-RF	ILMPFTC
1	64.77 ± 15.64	**75.23 ± 13.84**	70.00 ± 15.20	73.18 ± 12.77
2	75.26 ± 9.48	78.56 ± 13.05	**90.65 ± 8.89**	73.07 ± 11.74
3	68.70 ± 7.28	68.29 ± 13.46	**77.64 ± 9.66**	68.40 ± 12.76
4	67.92 ± 4.61	80.61 ± 5.37	**84.29 ± 7.02**	65.57 ± 8.38
5	72.18 ± 8.81	**72.61** ± 6.96	69.34 ± 8.10	72.57 ± 12.03
6	**70.13 ± 7.14**	67.00 ± 6.61	69.94 ± 5.40	64.94 ± 6.60
7	84.93 ± 3.36	**86.96** ± **2.90**	86.81 ± 5.27	79.42 ± 4.03
8	91.40 ± 3.67	82.15 ± 4.92	87.96 ± 2.90	**95.37** ± **2.10**
9	25.42 ± 2.89	**26.00** ± 4.37	24.70 ± 3.09	24.95 ± 4.41
10	68.52 ± 4.01	74.34 ± 3.63	**79.55 ± 2.28**	72.23 ± 3.34
11	90.83 ± 1.82	93.41 ± 1.38	**95.48 ± 1.75**	91.63 ± 1.33
12	90.14 ± 0.75	90.84 ± 1.17	**92.37** ± **0.71**	90.45 ± 0.50
Average rank	2.92	2.08	**2.00**	3.00

Table 18
Classification accuracy of embedding MFCCL in CM-CCL.

Dataset	CM-CCL-KNN	CM-CCL-DT	CM-CCL-RF	MFCCL
1	66.44 ± 17.97	75.23 ± 13.84	70.00 ± 15.20	**75.83 ± 11.72**
2	79.93 ± 9.00	86.08 ± 9.70	**91.90 ± 7.93**	78.79 ± 7.57
3	71.06 ± 7.73	69.76 ± 13.03	**78.16 ± 9.78**	68.40 ± 12.76
4	67.92 ± 4.61	80.28 ± 5.73	**84.62 ± 6.69**	70.92 ± 6.22
5	71.90 ± 7.57	71.91 ± 8.32	72.24 ± 8.89	**73.20 ± 10.65**
6	69.44 ± 6.90	67.53 ± 4.85	**70.29 ± 5.27**	64.94 ± 6.60
7	84.93 ± 3.36	86.67 ± 3.19	**86.81 ± 5.27**	79.42 ± 4.03
8	90.47 ± 3.55	82.01 ± 5.06	88.22 ± 2.97	**95.37** ± **2.10**
9	25.42 ± 2.89	26.00 ± 4.37	24.92 ± 3.15	**26.14** ± **4.30**
10	68.52 ± 4.01	74.34 ± 3.63	**79.55** ± **2.28**	72.23 ± 3.34
11	90.87 ± 1.48	94.03 ± 1.78	**95.58** ± **1.70**	91.44 ± 1.68
12	91.39 ± 0.67	89.99 ± 0.87	**92.63** ± **0.76**	90.45 ± 0.50
Average rank	3.08	2.58	**1.67**	2.67

model is replaced with MFCCL and ILMPFTC, with KNN, DT, and RF serving as classifiers. The experimental results are shown in Tables 17 and 18. It can be seen that after embedding ILMPFTC in the CM-CCL framework, its average rank in classification performance under the KNN, DT, and RF classifiers is 2.92, 2.08, and 2.00 respectively, exceeding its original rank of 3.00. For MFCCL, after embedding it in the CM-CCL framework, its average rank in classification performance under the KNN, DT, and RF classifiers is 3.08, 2.58, and 1.67 respectively, compared to its original rank of 2.67. The experimental results indicate that the proposed CM-CCL framework is scalable and capable of embedding existing CCL models to enhance the representation ability of concepts by leveraging multi-scale information.

5. Conclusion

How to effectively represent the knowledge contained in data has always been a meaningful question. Concept-cognitive learning provides a framework for knowledge representation that simulates the way humans perceive and understand things. Building on the foundation of existing CCL models, this paper has introduced multi-scale information into the process of learning concepts to enhance their representational capabilities. Specifically, a multi-scale data construction method based on the characteristics of CCL has been proposed, along with a multi-scale feature selection approach. On this basis, progressive concepts with stronger representational abilities are learned by fusing similar concepts. Besides, a corresponding collaborative mechanism has been designed to integrate these progressive concepts across multiple scales for the classification of unlabeled objects. The experimental results indicate that, on the one hand, the representation capability of concepts is stronger than that of data, and on the other hand, the proposed multi-scale concept-cognitive learning model outperforms the single-scale concept-cognitive learning model. Overall, the proposed CM-CCL has provided a knowledge representation paradigm that can accurately represent the valuable information in data.

This study broadens the scope of CCL, demonstrating that CM-CCL represents knowledge more comprehensively than single-scale CCL. Under the CM-CCL framework, there are still some issues worth further investigating.

(1) The multi-scale collaborative mechanism introduced in this paper serves as a method to integrate knowledge from multiple views. Therefore, it is a meaningful issue to investigate the application of this collaborative mechanism in multi-view or multi-modal data.
(2) CM-CCL fully exploits the intents of concepts for classification. It should be noted that the extents of concepts naturally provide a method for cross-scale information retrieval. Therefore, exploring the use of extents for multi-scale information fusion is worth investigating.

CRediT authorship contribution statement

Weihua Xu: Conceptualization, Funding acquisition, Investigation, Methodology, Project administration, Resources, Supervision, Validation; **Jinbo Wang:** Data curation, Formal analysis, Methodology, Software, Visualization, Writing – review & editing; **Qinghua Zhang:** Investigation, Data curation, Validation.

Data availability

Data will be made available on request.

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.

Acknowledgment

This work was supported in part by the National Natural Science Foundation of China under Grant 62376229 and in part by the Natural Science Foundation of Chongqing under Grant CSTB2023NSCQ-LZX0027.

References

[1] W. Wu, Y. Leung, J. Mi, Granular computing and knowledge reduction in formal contexts, IEEE Trans. Knowl. Data Eng. 21 (10) (2008) 1461–1474.
[2] C.J. Brainerd, Cognitive development and concept learning: an interpretative review, Psychol. Bull. 84 (5) (1977) 919.
[3] R. Wille, Restructuring lattice theory: an approach based on hierarchies of concepts, in: Proceedings of the NATO Advanced Study Institute, 1982, pp. 445–470.
[4] R. Bělohlávek, V. Sklenář, J. Zacpal, Crisply generated fuzzy concepts, in: International Conference on Formal Concept Analysis, Springer, 2005, pp. 269–284.
[5] J. Li, C. Mei, W. Xu, Y. Qian, Concept learning via granular computing: a cognitive viewpoint, Inf. Sci. 298 (2015) 447–467.
[6] Y. Mi, W. Liu, Y. Shi, J. Li, Semi-supervised concept learning by concept-cognitive learning and concept space, IEEE Trans. Knowl. Data Eng. 34 (5) (2020) 2429–2442.
[7] W. Xu, J. Pang, S. Luo, A novel cognitive system model and approach to transformation of information granules, Int. J. Approx. Reason. 55 (3) (2014) 853–866.
[8] W. Xu, D. Guo, J. Mi, Y. Qian, K. Zheng, W. Ding, Two-way concept-cognitive learning via concept movement viewpoint, IEEE Trans. Neural Netw. Learn. Syst. 34 (10) (2023) 6798–6812.
[9] W. Xu, D. Guo, Y. Qian, W. Ding, Two-way concept-cognitive learning method: a fuzzy-based progressive learning, IEEE Trans. Fuzzy Syst. 31 (6) (2022) 1885–1899.
[10] K. Yuan, W. Xu, W. Li, W. Ding, An incremental learning mechanism for object classification based on progressive fuzzy three-way concept, Inf. Sci. 584 (2022) 127–147.
[11] D. Guo, W. Xu, Y. Qian, W. Ding, Fuzzy-granular concept-cognitive learning via three-way decision: performance evaluation on dynamic knowledge discovery, IEEE Trans. Fuzzy Syst. 32 (3) (2023) 1409–1423.
[12] C. Zhang, E.C.C. Tsang, W. Xu, Y. Lin, L. Yang, Incremental concept-cognitive learning approach for concept classification oriented to weighted fuzzy concepts, Knowledge-Based Syst. 260 (2023) 110093.
[13] 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.

[14] J. Wu, E.C.C. Tsang, W. Xu, C. Zhang, L. Yang, Correlation concept-cognitive learning model for multi-label classification, Knowledge-Based Syst. 290 (2024) 111566.
[15] D. Guo, W. Xu, Y. Qian, W. Ding, M-FCCL: memory-based concept-cognitive learning for dynamic fuzzy data classification and knowledge fusion, Inf. Fusion 100 (2023) 101962.
[16] J. Wang, W. Xu, W. Ding, Y. Qian, Multiview fuzzy concept-cognitive learning with high-order information fusion of fuzzy attributes, IEEE Trans. Fuzzy Syst. 32 (12) (2024) 6965–6978.
[17] H. Zheng, M. Zhang, M. Gong, A.K. Qin, T. Liu, F. Jiang, Multi-scale hierarchical feature fusion network for change detection, Pattern Recognit. 161 (2025) 111266.
[18] I. Hussain, S. Tan, J. Huang, Few-shot based learning recaptured image detection with multi-scale feature fusion and attention, Pattern Recognit. 161 (2025) 111248.
[19] N. Ji, Y. Sun, F. Meng, L. Pang, Y. Tian, Variable multi-scale attention fusion network and adaptive correcting gradient optimization for multi-task learning, Pattern Recognit. 162 (2025) 111423.
[20] W. Wu, Y. Leung, Theory and applications of granular labelled partitions in multi-scale decision tables, Inf. Sci. 181 (18) (2011) 3878–3897.
[21] F. Li, B.Q. Hu, A new approach of optimal scale selection to multi-scale decision tables, Inf. Sci. 381 (2017) 193–208.
[22] Y. Yang, Q. Zhang, F. Zhao, Y. Cheng, Q. Xie, G. Wang, Optimal scale combination selection based on genetic algorithm in generalized multi-scale decision systems for classification, Inf. Sci. 693 (2025) 121685.
[23] L. Wang, W. Wu, Z. Xie, A. Tan, Optimal scale combination selection in generalized multi-scale hybrid decision systems, Inf. Sci. 689 (2025) 121429.
[24] J. Wang, W. Wu, A. Tan, Multi-granulation-based knowledge discovery in incomplete generalized multi-scale decision systems, Int. J. Mach. Learn. Cybern. 13 (12) (2022) 3963–3979.
[25] P. Zhang, D. Wang, Z. Yu, Y. Zhang, T. Jiang, T. Li, A multi-scale information fusion-based multiple correlations for unsupervised attribute selection, Inf. Fusion 106 (2024) 102276.
[26] N. Zhou, S. Liao, H. Chen, W. Ding, Y. Lu, Semisupervised feature selection with multiscale fuzzy information fusion: from both global and local perspectives, IEEE Trans. Fuzzy Syst. 33 (6) (2025) 1825–1839.
[27] Z. Wang, H. Chen, Z. Yuan, J. Wan, T. Li, Multiscale fuzzy entropy-based feature selection, IEEE Trans. Fuzzy Syst. 31 (9) (2023) 3248–3262.
[28] X. Zhang, J. Lin, Scalable data fusion via a scale-based hierarchical framework: adapting to multi-source and multi-scale scenarios, Inf. Fusion 114 (2025) 102694.
[29] T. Yin, H. Chen, Z. Wang, K. Liu, Z. Yuan, S.J. Horng, T. Li, Feature selection for multilabel classification with missing labels via multi-scale fusion fuzzy uncertainty measures, Pattern Recognit. 154 (2024) 110580.
[30] K. Yuan, D. Miao, W. Pedrycz, H. Zhang, L. Hu, Multigranularity data analysis with zentropy uncertainty measure for efficient and robust feature selection, IEEE Trans. Cybern. 55 (2) (2025) 740–752.
[31] W. Zhang, J. Ma, S. Fan, Variable threshold concept lattices, Inf. Sci. 177 (22) (2007) 4883–4892.
[32] Y. Mi, Y. Shi, J. Li, W. Liu, M. Yan, Fuzzy-based concept learning method: exploiting data with fuzzy conceptual clustering, IEEE Trans. Cybern. 52 (1) (2020) 582–593.
[33] X. Liang, Y. Qian, Q. Guo, H. Cheng, J. Liang, AF: an association-based fusion method for multi-modal classification, IEEE Trans. Pattern Anal. Mach. Intell. 44 (12) (2021) 9236–9254.
[34] M. Friedman, The use of ranks to avoid the assumption of normality implicit in the analysis of variance, J. Am. Stat. Assoc. 32 (200) (1937) 675–701.
[35] P.B. Nemenyi, Distribution-Free Multiple Comparisons, Princeton University, 1963.
[36] I. Guyon, A. Elisseeff, An introduction to variable and feature selection, J. Mach. Learn. Res. 3 (2003) 1157–1182.

Weihua Xu received the Ph.D. degree in mathematics from School of Sciences, Xi'n Jiaotong University, Xi'n, China, in 2007, and the M.S. degree in mathematics from School of Mathematics and Information Sciences, Guangxi University, Nanning, China, in 2004. He is currently a Professor with the College of Artificial Intelligence, Southwest University, Chongqing, China. He has published 5 monographs and more than 180 articles in international journals. His current research interests include granular computing, cognitive computing, and information fusion. Dr. Xu serves as a Senior Member of Chinese Association for Artificial Intelligence (CAAI). He also serves on the Associate Editor of *International Journal of Machine Learning and Cybernetics* and *Journal of Intelligent and Fuzzy Systems*.

Jinbo Wang received the M.S. degree in mathematics from the School of Information Engineering, Zhejiang Ocean University, Zhoushan, China, in 2023. He is currently working toward the Ph.D. degree in computer science and technology from the College of Artificial Intelligence, Southwest University, Chongqing, China. His current research interests include cognitive computing and granular computing.

Qinghua Zhang received the B.S. degree from Sichuan University, Chengdu, China, in 1998, the M.S. degree from the Chongqing University of Posts and Telecommunications, Chongqing, China, in 2003, and the Ph.D. degree from Southwest Jiaotong University, Chengdu, in 2010. He was with San Jose State University, San Jose, CA, USA, as a Visiting Scholar in 2015. Since 1998, he has been with the Chongqing University of Posts and Telecommunications, where he is currently a Professor and was appointed as the Vice-President. His research interests include rough sets, fuzzy sets, granular computing and uncertain information processing.