Incremental Incomplete Concept-Cognitive Learning Model: A Stochastic Strategy

Zhiming Liu, Jinhai Li, Xiao Zhang, and Xi-Zhao Wang, *Fellow, IEEE*

***Abstract*— Concept-cognitive learning is an emerging area of cognitive computing, which refers to continuously learning new knowledge by imitating the human cognition process. However, the existing research on concept-cognitive learning is still at the level of complete cognition as well as cognitive operators, which is far from the real cognition process. Meanwhile, the current classification algorithms based on concept-cognitive learning models (CCLMs) are not mature enough yet since their cognitive results highly depend on the cognition order of attributes. To address the above problems, this article presents a novel concept-cognitive learning method, namely, stochastic incremental incomplete concept-cognitive learning method (SI2CCLM), whose cognition process adopts a stochastic strategy that is independent of the order of attributes. Moreover, a new classification algorithm based on SI2CCLM is developed, and the analysis of the parameters and convergence of the algorithm is made. Finally, we show the cognitive effectiveness of SI2CCLM by comparing it with other concept-cognitive learning methods. In addition, the average accuracy of our model on 24 datasets is 82.02%, which is higher than the compared 20 classification algorithms, and the elapsed time of our model also has advantages.**

***Index Terms*— Classification, concept-cognitive learning, granular computing, incremental learning, stochastic incomplete concept.**

I. Introduction

COGNITIVE science is a discipline that uses scientific methods to study human mental world from the viewpoint of modern science, including many research fields such as psychology and computer science. Among these, cognitive computing is the core technical field in cognitive science and uses a computer system to simulate the human brain's cognition process [1], [2]. As a new computing paradigm, it reflects human cognition processes, including concept learning [3], [4] and brain thinking [5].

As the basic units of cognition, concepts are the foundation of human brain thinking [6]. As an emerging research direction, concept learning has attracted a large amount of attention from researchers. For example, the study of concept learning was made from different perspectives, such as granular computing [7], [8], [9], [10], [11], formal concept analysis [12], [13], [14], [15], and rough set [16], [17], [18], [19]. Learning concepts by recognizing necessary attributes and excluding unnecessary attributes is known as concept cognition. Yao [20] proposed a conceptual framework to explain concept cognition. Qiu et al. [21] presented a granular computing system to form different types of concepts. Kumar et al. [12] combined formal concept analysis with human brain's cognition to study concept cognition from a cognitive psychology perspective. Zhao et al. [22] studied concept cognition under incomplete information environment, and Fan et al. [23] put forward an attribute-oriented approach to concept cognition from a multilevel perspective. In addition, concept cognition was also combined with other domains, and some significant results have been achieved, such as three-way concept cognition [24], [25], [26], fuzzy bidirectional cognition [13], fuzzy incremental cognition [26], [27], [28], [29], concept cognition of causal asymmetric analysis [30], and multiattention concept-cognitive learning [31].

Zhang and Xu [32] put forward some concept-cognitive operators to form sufficient, necessary, and sufficient and necessary granules, making a study of concept-cognitive operators. On this basis, Xu et al. [33] further proposed a concept-cognitive learning method that converts all information granules into sufficient, necessary, and sufficient and necessary granules. Since it is often difficult to achieve exact cognition of concepts, Li et al. [34] presented an approximate concept-cognitive learning method based on the idea of upper and lower approximations. Considering that concept-cognitive learning is often affected by many factors such as time, space, and cost, Li et al. [35] incorporated the time factor into the concept-cognitive operator and gave an incremental concept-cognitive learning method under incomplete cognition.

In addition, inspired by the combination of granular computing and machine learning [36], Mi et al. [27], [37], [38] and Shi et al. [39], [40] combined concept cognition with machine learning and proposed fuzzy concept-cognitive learning classification algorithm [27], semisupervised concept-cognitive learning model (CCLM) [37], and incremental concept-cognitive learning classification algorithm [39]. To improve the efficiency of concept cognition, Mi et al. [38] and Shi et al. [40] further put forward parallel

Manuscript received 1 November 2022; revised 1 September 2023; accepted 13 November 2023. This work was supported by the National Natural Science Foundation of China under Grant 11971211 and Grant 12171388. *(Corresponding author: Jinhai Li.)*

Zhiming Liu and Jinhai Li are with the Data Science Research Center and the Faculty of Science, Kunming University of Science and Technology, Kunming 650500, China (e-mail: liuzhimingkust@163.com; jhlixjtu@163.com).

Xiao Zhang is with the Department of Applied Mathematics, Xi'an University of Technology, Xi'an 710048, China (e-mail: zhangxiao@xaut.edu.cn).

Xi-Zhao Wang is with the College of Computer Science and Software Engineering, Shenzhen University, Shenzhen 518060, China (e-mail: xizhaowang@ieee.org).

Color versions of one or more figures in this article are available at https://doi.org/10.1109/TNNLS.2023.3333537.

Digital Object Identifier 10.1109/TNNLS.2023.3333537

CCLMs. Meanwhile, Niu et al. [28], [41] explored a classification model based on granular concepts and dynamic granular rules, which was further extended to a fuzzy environment.

However, the existing studies on concept-cognitive learning still have some limitations. For example, the conversion of information granules into sufficient and necessary granules leads to some deviation between cognitive results and the target clue in [33]. In addition, there are many factors affecting the cognition process in real world. Li et al. [35] discussed the cognition step from the viewpoint of time, but it is unclear how to determine an appropriate cognition step and whether it can improve the cognitive accuracy. In fact, there are other factors affecting concept-cognitive learning except time, such as the factors of space and cost. In this article, the influences of time, space, and cost on concept-cognitive learning are uniformly viewed as priori knowledge. Generally speaking, different priori knowledge will lead to different cognitive results, but a better concept-cognitive learning method often guarantees that only changing the cognition order of attributes does not affect the final cognitive results. Thus, we propose stochastic incomplete concepts that are able to make the obtained cognitive results invariant when changing the cognition order of attributes.

On the other hand, in the application of concept-cognitive learning to classification task, some classification algorithms, such as CCLM [39], dynamic rule-based classification model (DRCM) [41], multiattention CCLM (MACLM) [31], and incremental learning mechanism based on progressive fuzzy three-way concept (ILMPFTC) [26], were developed for supervised learning. The drawbacks of these algorithms are apparent: 1) the CCLM algorithm involved high time complexity in calculating all granular concepts in the classification process; 2) the DRCM algorithm required the dataset to strictly meet the consistency condition, which is not suitable for the majority of real datasets; 3) the MACLM algorithm included too many parameters to achieve an effective concept clustering; and 4) the ILMPFTC algorithm did not consider the influence of priori knowledge on cognition, resulting in unsatisfactory cognitive results. In order to avoid these problems, we propose a stochastic incremental incomplete concept-cognitive learning method (SI2CCLM), which does not need to compute granular concepts, can make full use of priori knowledge in the cognition process, and is applicable for common datasets.

It is well known that stochasticity is undirected and can be characterized by multiple possibilities and fluctuations [42], [43], [44]. Human cognitive activities, based on natural, physiological, psychological, and social activities, have the stochastic characteristics. Meiser [45] analyzed the stochastic dependence of cognition process, and Warren et al. [46] pointed out that humans have biases in the random cognition process. In essence, cognition with a stochastic strategy is primarily about the randomness of recognizing attributes. Therefore, in order to reflect the randomness in human cognition in a quantitative way, we use a stochastic strategy by means of posterior probability to learn concepts, which is independent of the order of attributes.

Inspired by the above discussion, an incremental incomplete concept-cognitive learning method is proposed in this article. Specifically, the notion of a stochastic incomplete concept is given, and a stochastic incremental incomplete CCLM is discussed, which is further applied to the classification task. In addition, the analysis of parameters involved in the algorithms is made, how to determine and calculate them is investigated, and the convergence of the algorithm is proved. In the experiments, we show the cognitive effectiveness of the stochastic strategy in our method (SI2CCLM) by comparing it with the existing concept-cognitive learning methods. On the other hand, we evaluate the performance of SI2CCLM-based classification algorithm by comparing it with 16 machine learning classification algorithms and four concept-cognitive learning classification algorithms. It is shown that the average accuracy of SI2CCLM-based classification algorithm on 24 datasets is higher than those of the compared 20 classification algorithms, and our algorithm also has advantages in the running time.

The remainder of this article is organized as follows. Section II briefly introduces the preliminaries about concept-cognitive learning and formal decision context. Section III proposes the theory of SI2CCLM. In Section IV, we present the SI2CCLM-based classification algorithm, discuss the parameters setting, and prove its convergence. Some experiments are conducted in Section V to show the effectiveness of the proposed concept-cognitive learning method. A summary of the current work and some suggestions for future research are given in Section VI.

II. Preliminaries

In this section, we review some notions related to the concept-cognitive learning and formal decision context.

A. Concept-Cognitive Learning

Definition 1 [47]: A triple (G, M, I) is called a formal context, where $G = \{g_1, g_2, \ldots, g_m\}$ is a nonempty finite set of objects, $M = \{m_1, m_2, \ldots, m_n\}$ is a nonempty finite set of attributes, and I is a Boolean relation on the Cartesian product $G \times M$. $(g, m) = 1$ is called the object g possessing the attribute m, while $(g, m) = 0$ means the opposite. The complete concept-cognitive operators can be defined as follows:

$$\mathcal{L}(X) = \{m \in M | \forall g \in X, (g, m) = 1\} \tag{1}$$

$$\mathcal{H}(B) = \{g \in G | \forall m \in B, (g, m) = 1\} \tag{2}$$

where $X \subseteq G$ and $B \subseteq M$. $\mathcal{L}(X)$ is the set of the attributes common to all the objects in X, and $\mathcal{H}(B)$ denotes the set of the objects that have all the attributes in B.

Denote 2^G and 2^M as the power sets of G and M, respectively, and $\mathcal{L} : 2^G \to 2^M$ and $\mathcal{H} : 2^M \to 2^G$ are a pair of set-valued mappings.

Definition 2 [34]: $\mathcal{L}$ and $\mathcal{H}$ form a pair of complete concept-cognitive operators, if for $X_1, X_2 \subseteq G$ and $B_1, B_2 \subseteq M$, $\mathcal{L}$ and $\mathcal{H}$ satisfy the following properties.

1) $X_1 \subseteq X_2 \Rightarrow \mathcal{L}(X_2) \subseteq \mathcal{L}(X_1)$.
2) $B_1 \subseteq B_2 \Rightarrow \mathcal{H}(B_2) \subseteq \mathcal{H}(B_1)$.
3) $X \subseteq \mathcal{HL}(X), B \subseteq \mathcal{LH}(B)$.

Also, $\mathcal{L}$ is called the object–attribute operator and $\mathcal{H}$ is called the attribute–object operator.

For convenience, hereinafter, we refer to $\mathcal{L}$ as the complete cognitive operator, and concepts generated by the operators $\mathcal{L}$ and $\mathcal{H}$ are called complete cognitive concepts.

Definition 3 [47]: Let (G, M, I) be a formal context. For $X \subseteq G$ and $B \subseteq M$, if $\mathcal{L}(X) = B$ and $\mathcal{H}(B) = X$, then the pair (X, B) is said to be a formal concept, where X is the extent of the concept and B is the intent of the concept.

Definition 4: Let (G, M, I) be a formal context. For $X_1, X_2, \ldots, X_k \subseteq G$ and $B_1, B_2, \ldots, B_k \subseteq M$, if $\mathcal{C}_1 = (X_1, B_1), \mathcal{C}_2 = (X_2, B_2), \ldots, \mathcal{C}_k = (X_k, B_k)$ are k formal concepts and they are further used to do cognition of new concepts, then $\mathcal{C}_1, \mathcal{C}_2, \ldots, \mathcal{C}_k$ are called the priori knowledge with its length being k.

A priori knowledge with its length being k can be generated randomly in the experiments to simulate what has already been learned in the cognition process.

B. Formal Decision Context

Definition 5 [39]: A quintuple (G, M, I, D, J) is said to be a regular formal decision context, where $I \subseteq G \times M$ and $J \subseteq G \times D$, and for any $d_1, d_2 \in D$, $\mathcal{H}(d_1) \cap \mathcal{H}(d_2) = \emptyset$. (G, M, I) and (G, D, J) are called the conditional formal context and decision formal context, respectively.

That is, there exists a unique label $d\,(d \in D)$ for each object $g\ (g \in G)$ in a regular formal decision context.

Definition 6 [39]: Let $\mathcal{K} = (G, M, I, D, J)$ be a regular formal decision context, and $D = D_1 \cup D_2 \cup \cdots \cup D_l (D_i \cap D_j = \emptyset)$. Then, $\mathcal{K}^{D_i} = (G_i, M_i, I_i, D_i, J_i)$ is called the decision subcontext of $\mathcal{K}$ under the label of D_i, and we have $\mathcal{K} = \bigcup_{i \in L} \mathcal{K}^{D_i}$, where $L = \{1, 2, \ldots, l\}$ is a set of labels.

From Definition 6, a regular formal decision context can be decomposed into l decision subcontexts by l labels. For brevity, the concept mapping under the label of D_i is denoted as $\mathcal{H}^{D_i}$.

III. Stochastic Incremental Incomplete Concept-Cognitive Learning Method

In human cognition, sometimes changing the order of cognition of things does not affect the final cognitive results. However, the existing concept-cognitive learning methods are not able to deal with this problem since they are sensitive to the cognition order of things. In this section, we define a stochastic incomplete concept and propose an SI2CCLM to learn the concept from a given clue, which is not sensitive to the cognition order of attributes.

A. Stochastic Incremental Incomplete Concept-Cognitive Learning Process

Concept-cognitive learning is a process of learning concepts from given clues, which can be sets of objects or sets of attributes. Generally speaking, there may be many concepts that can be learned from the same clue, but there is only one concept whose extent contains the clue with the least number of objects. When a clue is an extent of a concept, the cognition is to learn the concept with the clue being its extent; when a clue is not an extent of a concept, then it is interesting to find a minimal concept whose extent contains the clue. In any case, our goal is to learn the best matching concept through a clue.

The key part of the process of learning a concept is to recognize the attributes associated with the clue. Cognition of attributes is often a random process and it is influenced by multiple factors such as time, space, and cost. Thus, it is important to measure the possibility of cognition of attributes for the purpose of realizing incremental learning of attributes. In the following, we first give a probability to each attribute being recognized preferentially for developing an SI2CCLM. The event of recognizing an attribute from a clue X_0 is considered as a random variable ξ, and $P(\xi = m_i, X_0)$ represents the probability of an attribute m_i being recognized from the clue X_0 at a certain time, which is abbreviated as $P(m_i, X_0)$ below for simplicity.

Definition 7: Let (G, M, I) be a formal context, $X_0 \subseteq G$, $M = \{m_i | i = 1, 2, \ldots, n\}$, and $\mathcal{H}$ be an attribute–object operator. When cognition is performed on the clue X_0, the probability of an attribute m_i being preferentially recognized is defined as

$$P(m_i, X_0) = \exp\left(-\frac{\cos^{-1}(\mathcal{H}(m_i), X_0)}{2\sigma^2}\right) \tag{3}$$

where σ is a parameter. When $\cos(\mathcal{H}(m_i), X_0) = 0$, let $P(m_i, X_0) = 0$.

In fact, the right-hand side of (3) is the combination of the Gaussian kernel function [48] and the cosine similarity [49], which is used to measure the similarity degree between the objects of m_i and the clue X_0.

In many random phenomena, $P(m_i, X_0)$ having the largest value does not completely mean that the attribute m_i must be recognized first, but it indeed represents that the attribute m_i has the highest chance being recognized. In (3), although the probability of each attribute being preferentially recognized is defined for a given clue, we need to normalize all the probabilities to realize the random selection of attributes based on roulette principle, i.e.,

$$P_{\text{new}}(m_i, X_0) = \frac{P(m_i, X_0)}{\sum_{i=1}^{n} P(m_i, X_0)}. \tag{4}$$

That is, the stochastic cognition of attributes is converted to the random selection of attributes based on the probabilities. In the following, roulette will use these probabilities $P_{\text{new}}(m_i, X_0)$ $(m_i \in M)$ to realize the random selection of attributes. For brevity, $P_{\text{new}}(m_i, X_0)$ is also written as $P(m_i, X_0)$ if causing no confusion.

Once an attribute is selected, the first round of cognition is completed. However, it still needs to judge whether the identified attribute is correct, i.e., the selected attribute must be a substantial attribute of the clue X_0.

Definition 8: Let (G, M, I) be a formal context, $\mathcal{H}$ be an attribute–object operator, $X_0 \subseteq G$ be a clue, and m_i be the attribute selected in the process of cognition. If $X_0 \subseteq \mathcal{H}(m_i)$ is satisfied, then the cognition of m_i is said to be valid.

When the cognition is valid, the selected attribute is retained and we continue doing the next cognition. Otherwise, the random selection needs to be performed repeatedly until an attribute is successfully selected. Only, in this case, can we achieve the first round of cognition. However, if the selection is repeated many times, but we still cannot get the required attribute, then we say that no attribute can be recognized for the given clue.

Note that the previous selection of attributes will more or less cause some impact on the probability of selecting the remaining attributes since there is a dependency between them. Therefore, we continue to introduce a posteriori probability to measure the probability of selecting the remaining attributes after one or more attributes have been selected. The conditional probability $P(m_j|m_i, X_0)$ can represent the probability of selecting the attribute m_j for the clue X_0 after the selection of the attribute m_i has been completed.

Definition 9: Let $\{m_i\}_{i=1}^n$ be the set of attributes of (G, M, I). Suppose that the attributes $m_1, m_2, \ldots, m_l$ $(l < n)$ have been recognized in the previous cognition. Then, the posterior probability of selecting the attribute m_k is defined as

$$P(m_k|m_1, \ldots, m_l, X_0) = \exp\left(-\frac{\cos^{-1}(\mathcal{H}(m_k), X_0) \cdot \prod_{i=1}^{l} \cos^{-1}(\mathcal{H}(m_i), \mathcal{H}(m_k))}{2\sigma^2}\right) \tag{5}$$

where σ is a parameter. When $\cos(\mathcal{H}(m_k), X_0) = 0$ or $\cos(\mathcal{H}(m_i), \mathcal{H}(m_k)) = 0$, let $P(m_k|m_1, \ldots, m_l, X_0) = 0$.

According to Definition 9, if both the attributes m_i and m_k share the same set of objects, we have $P(m_k|m_i, X_0) = P(m_k, X_0)$; otherwise, $P(m_k|m_i, X_0) < P(m_k, X_0)$. By Definitions 7 and 9, we can calculate the probabilities of all the attributes being recognized, which help us gradually select all the necessary attributes and, at the same time, excludes all the unnecessary attributes associated with the clue X_0 until the cognition is finished.

Property 1: For any integer $h \in [1, l)$, let $P_1 = P(m_k|m_1, \ldots, m_h, X_0)$ and $P_2 = P(m_k|m_{h+1}, \ldots, m_l, X_0)$. Then, the following equation holds:

$$P(m_k|m_1, \ldots, m_l, X_0) = \exp\left(\frac{1}{A} \cdot \ln P_1 \cdot \ln P_2\right) \tag{6}$$

where $A = -((\cos^{-1}(\mathcal{H}(m_k), X_0))/2\sigma^2)$ is free of the information of the selected attributes.

Proof: By taking the logarithm function $\ln(\cdot)$ of both sides of (5), we have

$$\begin{aligned}
&\ln P(m_k|m_1, \ldots, m_l, X_0) \\
&= A \cdot \prod_{i=1}^{l} \cos^{-1}(\mathcal{H}(m_i), \mathcal{H}(m_k)) \\
&= A \cdot \prod_{i=1}^{h} \cos^{-1}(\mathcal{H}(m_i), \mathcal{H}(m_k)) \prod_{j=h+1}^{l} \cos^{-1}(\mathcal{H}(m_j), \mathcal{H}(m_k)) \\
&= \frac{1}{A} \cdot \ln P(m_k|m_1, \ldots, m_h, X_0) \cdot \ln P(m_k|m_{h+1}, \ldots, m_l, X_0).
\end{aligned}$$

Furthermore, by taking the exponential function $\exp(\cdot)$ of both sides of the above equation, we obtain

$$P(m_k|m_1, \ldots, m_l, X_0) = \exp\left(\frac{1}{A} \cdot \ln P_1 \cdot \ln P_2\right)$$

that is, the conclusion is at hand. ■

By Property 1, the probability $P(m_k|m_1, \ldots, m_l, X_0)$ of a newly added attribute can be obtained by integrating the local subprobabilities P_1 and P_2, which means that the calculation of (5) can be accelerated by parallel computing.

B. Stochastic Incomplete Concept

Given a clue X_0, suppose that all the cognitive attributes obtained are recorded as B_p. Then, the notion of a stochastic incomplete concept is defined as follows.

Definition 10: Let (G, M, I) be a formal context, $\mathcal{H}$ be an attribute–object operator, and X_0 be a clue. The set of all the attributes obtained by the cognition process is denoted by B_p and $X_p = \mathcal{H}(B_p)$. Then, we call (X_p, B_p) a stochastic incomplete concept.

Property 2: Let (X_p, B_p) be the stochastic incomplete concept of a given clue X_0 obtained by the cognitive process. Then, $X_0 \subseteq X_p$ and $B_p \subseteq \mathcal{L}(X_0)$.

Proof: The proof is immediate from Definitions 1, 2, and 10. ■

In fact, the cognition process will be terminated when all the attributes associated with the clue have been recognized. In this case, there is no attribute in the candidate set whose objects include the given clue X_0.

Property 3: Let (X_p, B_p) be the stochastic incomplete concept of a given clue X_0 obtained by the cognitive process. Then, there is no attribute in the candidate set whose objects include the given clue X_0 if and only if $(X_p, B_p) = (\mathcal{HL}(X_0), \mathcal{L}(X_0))$.

Proof: $(X_p, B_p) = (\mathcal{HL}(X_0), \mathcal{L}(X_0))$ is equivalent to $B_p = \mathcal{L}(X_0)$ since a concept can uniquely be determined by its intent. What is more, $B_p = \mathcal{L}(X_0)$ is equivalent to the assumption that there is no attribute in the candidate set whose objects include the given clue X_0. To sum up, the proof is completed. ■

From Definition 10 and Property 3, it can be seen that the stochastic incomplete concept is different from the formal concept because the process of recognizing attributes through the clue is an incomplete cognition process. However, when $B_p = \mathcal{L}(X_0)$, the stochastic incomplete concept becomes a formal concept. In this case, the stochastic incremental incomplete concept cognition achieves complete cognition.

Based on the above discussion, to achieve a complete cognition of a given clue X_0, we need to recognize all the substantial attributes of the given clue in the cognition process. Motivated by this, the stochastic incremental incomplete concept-cognitive learning with the clue X_0 is summarized in Algorithm 1. For brevity, we abbreviate the algorithm as SI2CCLM.

To make the algorithm easier to understand, we further illustrate it in the scene of classifying objects into a cluster based on a concept, that is, how to learn an appropriate concept, which is able to classify the objects into a cluster and keep the cluster they belong to correct. To make it simple, suppose that the task is to classify all the objects in X_0. Although there may be more than one concept achieving this task, it is important to find the best one that can not only classify the objects into a cluster but also exclude other objects not belonging to the cluster as much as possible. In fact, finding such a concept is equivalent to recognizing all the substantial attributes of X_0. Generally speaking, it is difficult to recognize all the substantial attributes for a given clue since the recognized attributes will more or less disturb the selection of other attributes. This requests that the process of recognizing the attributes can be proceeded until the task is achieved no matter what the order of the attributes being recognized is. Our algorithm can satisfy this request since selecting an attribute each time in the cognition process is stochastic by means of the probability (i.e., the similarity between the candidate attribute and the given clue). In fact, the stochasticity used in our algorithm can guarantee that the output is as globally optimal as possible and it is just the target concept of X_0, which can classify the objects correctly.

Now, we continue to analyze the time complexity of Algorithm 1. The time complexity of Steps 4–12 and 19–32 is $O(|G||M|)$, while that of Steps 3 and 18 is $O(|G||M|)$. Thus, the time complexity of Steps 3–12 is $O(|G||M|)$, and that of Steps 17–33 is $O(q|G||M|)$, where q is the cardinality of substantial attributes of the clue X_0. Therefore, the total time complexity of our algorithm is $O(q|G||M|)$ in the worst case.

The above only discusses the stochastic incremental incomplete concept-cognitive learning for the clue being a set of objects. However, it can also be a set of attributes. In this case, we can use the concept-cognitive operator $\mathcal{H}$ to transform attributes into objects before concept cognition, so we do not elaborate on this case in this article.

IV. Stochastic Incremental Incomplete Concept-Cognitive Learning Classification Model

In this section, we further apply stochastic incomplete concepts to dealing with the classification problem and put forward a classification model based on SI2CCLM. Specifically, we discuss the concept prediction of SI2CCLM-based classification model, analyze the parameters involved in our model, and prove the convergence of our algorithm.

A. Concept Prediction Process of SI2CCLM-Based Classification Model

How to predict the labels of instances becomes a critical problem when the SI2CCLM is applied to the classification task. In the experiments, the dataset will randomly be divided into a training set G_{train} and a testing set G_{test}. For every predictive instance $g_s \in G_{\text{test}}$, an object set can be obtained in each decision subcontext associated with the label D_i ($i \in \{1, 2, \ldots, l\}$) by using the concept-cognitive operator $\mathcal{H}^{D_i}$. These obtained object sets are considered as cognitive clues. By Algorithm 1, l stochastic incomplete concepts can be obtained based on these cognitive clues, and then, we compute similarity degrees between the predictive instance g_s and l stochastic incomplete concepts. Furthermore, we are able to determine the greatest similarity degree and use it to complete the assignment of the predictive label of the instance g_s.

According to the above discussion, it needs to clarify the similarity degree between the predictive instance and stochastic incomplete concepts.

Algorithm 1 Stochastic Incremental Incomplete Concept-Cognitive Learning From a Clue X_0

1: **Input:** A formal context (G, M, I), the clue X_0, attribute-object operator $\mathcal{H}$ and σ.
2: **Output:** A stochastic incomplete concept (X_p, B_p).
3: Calculate the probabilities $P(m_i, X_0)$ $(m_i \in M)$ of attributes being selected;
4: **for** $k = 1 : |M|$ **do**
5:   Roulette is used to select an attribute m_i randomly;
6:   **if** $X_0 \subseteq \mathcal{H}(m_i)$ **then**
7:     $B_p = \{m_i\}$;
8:     **break**
9:   **else**
10:     **continue**
11:   **end if**
12: **end for**
13: **if** $B_p = \emptyset$ **then**
14:   $X_p = U$, and go to Step 32;
15: **end if**
16: Let $q = 1$;
17: **while** $|B_p| = q$ **do**
18:   Calculate the conditional probabilities of the remaining attributes other than the attributes $m_1, m_2, \ldots, m_q$;
19:   **for** $h = 1 : |M|$ **do**
20:     Roulette is used to select an attribute randomly;
21:     **if** $X_0 \subseteq \mathcal{H}(m_{q+1})$ **then**
22:       $q = q + 1$;
23:       $B_p = \{m_1, m_2, \ldots, m_q, m_{q+1}\}$;
24:       **break**;
25:     **else**
26:       **continue**;
27:     **end if**
28:     **if** $h = |M|$ **then**
29:       $q = q + 1$;
30:       $B_p = \{m_1, m_2, \ldots, m_q\}$;
31:     **end if**
32:   **end for**
33: **end while**
34: $X_p = \mathcal{H}(B_p)$;
35: **Return** (X_p, B_p).

Definition 11: Let $\mathcal{K}^{D_i} = (G_i, M_i, I_i, D_i, J_i)$ be a regular decision subcontext under the label D_i, where $i \in L$ ($L = \{1, 2, \ldots, l\}$). Suppose that (X_{p_i}, B_{p_i}) is a stochastic incomplete concept generated by the predictive instance g_s ($g_s \in G_{\text{test}}$) under $\mathcal{K}^{D_i}$. We use a vector $\vec{g_s}$ to represent the attributes' information possessed by g_s. Then, the similarity degree between the predictive instance g_s and the stochastic incomplete concept (X_{p_i}, B_{p_i}) is defined as follows:

$$\gamma_s^i = \frac{|\vec{g_s} \cap B_{p_i}|}{|\vec{g_s} \cap B_{p_i}| + 2(\alpha|\vec{g_s} - B_{p_i}| + (1-\alpha)|B_{p_i} - \vec{g_s}|)} \tag{7}$$

where $\alpha \in [0, 1]$.

For simplicity, (7) is called attribute similarity degree and it is abbreviated as AS degree. For the predictive instance g_s, we can obtain a vector $(\gamma_s^1, \gamma_s^2, \ldots, \gamma_s^l)$ for all the l labels. On this basis, the maximum of similarity degrees can be obtained, namely, $\hat{\gamma}_s = \max_{i \in L}\{\gamma_s^i\}$, where $L = \{1, 2, \ldots, l\}$. Then, the label of the stochastic incomplete concept with the maximum value $\hat{\gamma}_s$ is considered as the predicted label of g_s if the maximum value is unique. In this way, we can successfully assign labels to all predictive instances in the testing set.

In the following, we discuss the special case: the maximum similarity degrees between the predictive instance and the stochastic incomplete concepts are not unique. In this case, we have to further distinguish them from each other.

1) $\exists\ i, j, \ldots, k \in L,\ s.t.\ \hat{\gamma}_s = \gamma_s^i = \gamma_s^j = \cdots = \gamma_s^k \neq 0$.
2) $\gamma_s^1 = \gamma_s^2 = \cdots = \gamma_s^l = 0$.

Definition 12: Let $\mathcal{K}^{D_i} = (G_i, M_i, I_i, D_i, J_i)$ be a regular decision subcontext under the label D_i, where $i \in L$ ($L = \{1, 2, \ldots, l\}$). Suppose that $(X_{p_s}^1, B_{p_s}^1), (X_{p_s}^2, B_{p_s}^2), \ldots, (X_{p_s}^l, B_{p_s}^l)$ are the l stochastic incomplete concepts generated by g_s under l labels. If there exist $i, j, \ldots, k \in L$ such that $\hat{\gamma}_s = \gamma_s^i = \gamma_s^j = \cdots = \gamma_s^k \neq 0$, then the label of g_s is denoted as

$$l_{g_s} = l_{\max} \in \{i, j, \ldots, k\}$$

where $|X_{p_s}^{l_{\max}}| = \max(|X_{p_s}^i|, |X_{p_s}^j|, \ldots, |X_{p_s}^k|)$. When $l_{\max}$ is not unique, we choose one of them randomly as the label of g_s.

For the other case, if the similarity degrees between the predictive instance g_s and the stochastic incomplete concepts are 0, it means that the instance has no matching stochastic incomplete concept in each decision subcontext. To handle this problem, the similarity degrees between the predictive instance g_s and each instance of the training set are calculated, and the label of the instance in the training set with the highest similarity degree is assigned to the predictive instance.

To sum up, no matter which case happens, we are able to complete the concept prediction task for all the predictive instances in the testing set.

B. Overall Procedure of SI2CCLM-Based Classification Model

Based on the previous discussion, Fig. 1 shows the overall procedure of SI2CCLM-based classification model, which includes three stages: 1) concept mapping; 2) stochastic concept cognition; and 3) concept prediction. Assume that the training set in Stage 1 is divided into three classes according to three labels. First, for any predictive instance g_s in the testing set, concept mapping is done in three decision subcontexts under three labels, and then, three object sets can be obtained and viewed as the input of Stage 2. Second, apply Algorithm 1 to realizing stochastic incremental incomplete concept-cognitive learning of the three clues, and stochastic incomplete concepts $(X_{p_s}^i, B_{p_s}^i)$ are obtained. Finally, we compute similarity degrees between the predictive instance g_s and three stochastic incomplete concepts and use them to assign the most suitable category label to the instance g_s. The detailed process is described in Algorithm 2.

Now, we continue to analyze the time complexity of Algorithm 2. According to (7), the time complexity of Step 10 is $O(|M|)$. Note that Algorithm 2 will call Algorithm 1 in Step 9 to calculate a stochastic incomplete concept. Thus, the time complexity of Steps 4–12 is $O(q|G_{\text{train}}||M||D|)$,

Algorithm 2 Prediction Algorithm Based on SI2CCLM

1: **Input:** A regular decision context (G, M, I, D, J), training set G_{train}, and testing set G_{test}.
2: **Output:** The classification accuracy $Accu$.
3: **for** each $g_s \in G_{test}$ **do**
4: **for** each label D_i **do**
5: $X_s^i = \mathcal{H}^{D_i}(\vec{g_s})$;
6: **if** $X_s^i = \emptyset$ **then**
7: $\gamma_s^i = 0$;
8: **else**
9: Apply Algorithm 1 to concept cognition with the clue X_s^i, and get the stochastic incomplete concept $(X_{p_s}^i, B_{p_s}^i)$;
10: Compute the AS degree γ_s^i between the instance g_s and $B_{p_s}^i$ by Eq. (7);
11: **end if**
12: **end for**
13: Find the label $l_{\max}$ of the stochastic incomplete concept whose AS degree with g_s is the maximum;
14: **if** the maximum AS degree is 0, **then**
15: Find the most similar object to g_s from the training set G_{train} and take its label l_{g_s} as the predictive label of g_s;
16: **else if** $l_{\max}$ is unique **then**
17: $l_{g_s} = l_{\max}$;
18: **else**
19: $|X_{p_s}^j| = \max(|X_{p_s}^1|, |X_{p_s}^2|, \ldots, |X_{p_s}^k|)$;
20: $l_{g_s} = j$;
21: **end if**
22: **end for**
23: Let L_c be the number of all correctly predicted labels, and $Accu = L_c/|G_{test}|$;
24: **Return** $Accu$.

where q is the maximum cardinality of the clues of predictive instances. Obviously, the time complexity of Step 15 is $O(|G_{\text{train}}||M|)$, so is that of Steps 13–21. As a result, Steps 3–22 take $O(q|G_{\text{train}}||G_{\text{test}}||M||D|)$ time. It should be pointed out that before embarking on Algorithm 2, we need to call Algorithm 3 [see Section IV-C for details, and its time complexity is $O(|G_{\text{train}}||M|)$] to determine the value of the parameter σ. To sum up, the total time complexity of Algorithm 2 is $O(q|G_{\text{train}}||G_{\text{test}}||M||D|)$ in the worst case.

Note that the existing concept-cognitive learning classification models match predictive instances by exhausting all the granular concepts under each label to achieve the label prediction of instances. However, sometimes it is not enough for granular concepts to achieve the best match for the predictive instances, that is, it often happens that the granular concepts of each label do not cover enough required information and one has to find the next best match or even a reluctant match for the predictive instances, resulting in an inaccurate prediction. Our algorithm can effectively avoid such cases since the stochastic incomplete concepts of each label contain more information than granular concepts, and it matches the predictive instances with more accurate concepts. Thus, the classification performance of SI2CCLM-based classification

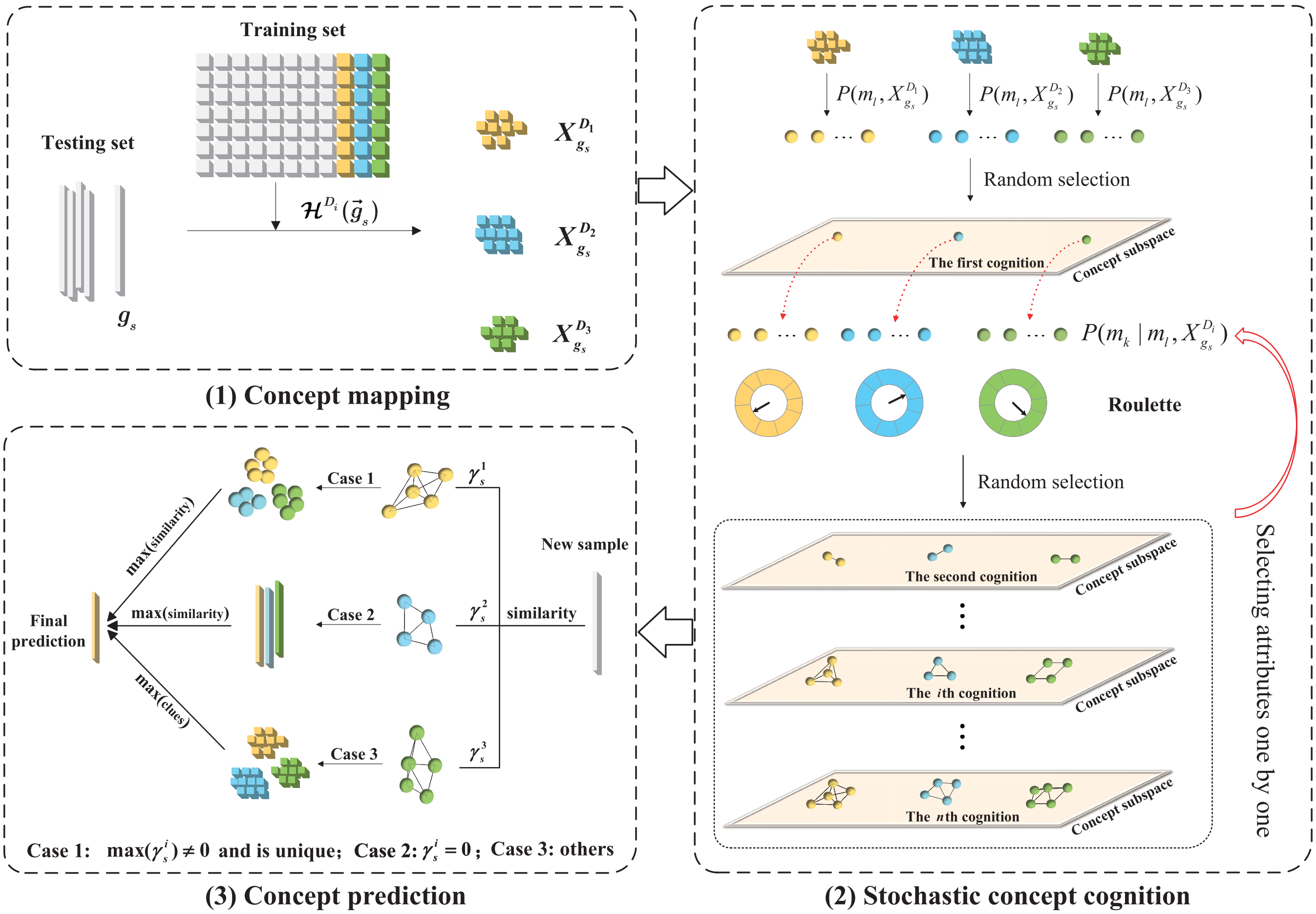

Fig. 1. Diagram of the overall procedure of SI2CCLM-based classification model.

model is expected to be better than those of the existing concept-cognitive learning classification models.

On the other hand, our algorithm is to find the best match from the obtained stochastic incomplete concepts whose number equals that of the labels, while the existing concept-cognitive learning methods are to find the best match, the next best match, or even a reluctant match from the granular concepts of the training set whose number is often much greater than that of the labels. Thus, our algorithm can reduce the running time in achieving the classification task.

C. Analysis of Parameters

Note that there are two parameters α and σ in Algorithm 2. The first parameter α was appeared in (7). Thus, the change of α will lead to an update of the value of (7), and consequently, the accuracy of SI2CCLM-based classification model will also be changed. In this article, we will determine an appropriate value for the parameter α by a reasonable experimental analysis, and the details are left in Section V.

The second parameter σ was brought in the process of stochastic incremental incomplete concept-cognitive learning. According to Algorithm 1, there is no doubt that the determination of the parameter σ plays an important role in developing a reasonable stochastic concept-cognitive learning process. Moreover, it can be known from (3) and (5) that different values of the parameter σ reflect the different probabilities of preferentially recognizing an attribute. Thus, how to find an appropriate value for the parameter σ is crucial to the improvement of the classification performance.

In what follows, we construct an objective function $Z(\sigma)$ to optimize the value of the parameter σ such that the probabilities between $\mathcal{H}(m_i)$ $(i = 1, 2, \ldots, n)$ and X_0 are as large as possible, while the probabilities between $\mathcal{H}(M - \{m_i\})$ $(i = 1, 2, \ldots, n)$ and the clue are as small as possible.

It is worth noting that the objects in the same decision subcontext cannot be distinguished from each other by only using their original labels that are the same. Thus, we need to find additional supervised information to obtain the differences between the attributes and the given clue.

Definition 13: Let (G, M, I) be a formal context and $\mathcal{C} = (X, B)$ be the priori knowledge with its length being 1. The set of all attributes in B is denoted as $\{m_i\}_{i \in \{k_1, k_2, \ldots, k_p\}}$, and the set of all attributes in $M - B$ is denoted as $\{m_j\}_{j \in \{t_1, t_2, \ldots, t_{n-p}\}}$. Then, we say that the label of m_i is 1 and that of m_j is -1, which are denoted as $y_{m_i} = 1$ and $y_{m_j} = -1$, respectively.

Essentially, the objective of the proposed stochastic incremental incomplete CCLM is to provide a clustering of the attributes of a given clue X_0 such that attributes in the same class are as similar as possible, while attributes in the different class are as dissimilar as possible. Motivated by the work in

[50], we construct the following objective function $Z(\sigma)$ to achieve the goal of parameter optimization:

$$Z(\sigma) = \sum_{i=1}^{n}\sum_{j=1}^{i-1} P(m_i, X_0)\cdot y_{m_i}\cdot y_{m_j} \tag{8}$$

where y_{m_i} and y_{m_j} represent the labels of the attributes m_i and m_j, respectively. It is sufficient to make $Z(\sigma)$ as large as possible to meet the above clustering requirement. In this way, the problem of parameter determination can easily be transformed into an optimization problem of finding the maximum value of the below function

$$\hat{\sigma} = \arg\max_{\sigma} Z(\sigma). \tag{9}$$

Theorem 1: When $\sigma = (1/\sqrt{-2c})$, we have $Z(\sigma) = Z_{\max}$, where

$$c = -\frac{\sum_{i=1}^{n}\sum_{j=1}^{i-1}\cos^{-1}(\mathcal{H}(m_i), X_0)\cdot y_{m_i}\cdot y_{m_j}}{\sum_{i=1}^{n}\sum_{j=1}^{i-1}\cos^{-2}(\mathcal{H}(m_i), X_0)\cdot y_{m_i}\cdot y_{m_j}}. \tag{10}$$

Proof: For conciseness, let

$$\lambda_i = \cos^{-1}(\mathcal{H}(m_i), X_0), \quad c = -\frac{1}{2\sigma^2}. \tag{11}$$

By substituting (3) and (11) into (8), it follows:

$$Z(\sigma) = Z(c) = \sum_{i=1}^{n}\sum_{j=1}^{i-1} e^{c\lambda_i}\cdot y_{m_i}\cdot y_{m_j}. \tag{12}$$

Expanding $e^{c\lambda_i}$ by using the Taylor expansion, we have

$$e^{c\lambda_i} \approx 1 + c\lambda_i + \frac{1}{2!}(c\lambda_i)^2. \tag{13}$$

By substituting (13) into (12), we obtain

$$Z(c) = \sum_{i=1}^{n}\sum_{j=1}^{i-1} y_{m_i}\cdot y_{m_j} + \left(\sum_{i=1}^{n}\sum_{j=1}^{i-1}\lambda_i\cdot y_{m_i}\cdot y_{m_j}\right)\cdot c + \left(\sum_{i=1}^{n}\sum_{j=1}^{i-1}\frac{1}{2!}\lambda_i^2\cdot y_{m_i}\cdot y_{m_j}\right)\cdot c^2. \tag{14}$$

Obviously, this is a quadratic function only related to c. In most cases, the number of attributes owned by the concept is smaller than the number of attributes not owned by the concept in a formal context, which usually satisfies

$$\sum_{i=1}^{n}\sum_{j=1}^{i-1}\frac{1}{2!}\lambda_i^2\cdot y_{m_i}\cdot y_{m_j} < 0.$$

Thus, the maximum value of $Z(c)$ can be reached when

$$c = -\frac{\sum_{i=1}^{n}\sum_{j=1}^{i-1}\lambda_i\cdot y_{m_i}\cdot y_{m_j}}{\sum_{i=1}^{n}\sum_{j=1}^{i-1}\lambda_i^2\cdot y_{m_i}\cdot y_{m_j}}. \tag{15}$$

That is, the proof of Theorem 1 is completed. ■

According to Theorem 1, we can obtain the optimal value of the parameter

$$\hat{\sigma} = \sqrt{\frac{\sum_{i=1}^{n}\sum_{j=1}^{i-1}\cos^{-2}(\mathcal{H}(m_i), X_0)\cdot y_{m_i}\cdot y_{m_j}}{2\sum_{i=1}^{n}\sum_{j=1}^{i-1}\cos^{-1}(\mathcal{H}(m_i), X_0)\cdot y_{m_i}\cdot y_{m_j}}}. \tag{16}$$

Although priori knowledge with its length being 1 (i.e., a concept) can be used to calculate one value of σ, the priori knowledge will be repeated k times in the subsequent experiments to reduce the impact of the randomness of priori knowledge. When the length of the priori knowledge is k (i.e., k concepts), σ is averaged over them. Usually, the values of σ are different when the datasets used to train it are different, which reflects the learning abilities of different datasets. The detailed procedure of calculating σ is given in Algorithm 3 whose time complexity is analyzed as follows.

The time complexity of calculating $\cos^{-1}(\mathcal{H}(m_i), X_0)$ and $\cos^{-2}(\mathcal{H}(m_i), X_0)$ for any $i \in \{1, 2, \ldots, n\}$ is $O(|G|)$ so that of $\sum_{i=1}^{n}\cos^{-1}(\mathcal{H}(m_i), X_0)\cdot y_{m_i}$ and $\sum_{i=1}^{n}\cos^{-2}(\mathcal{H}(m_i), X_0)\cdot y_{m_i}$ is $O(|G||M|)$. According to (16), the calculation of the parameter σ_r can be represented as

$$\sigma_r = \sqrt{\frac{\sum_{i=1}^{n}\cos^{-2}(\mathcal{H}(m_i), X_0)\cdot y_{m_i}\sum_{j=1}^{i-1} y_{m_j}}{2\sum_{i=1}^{n}\cos^{-1}(\mathcal{H}(m_i), X_0)\cdot y_{m_i}\sum_{j=1}^{i-1} y_{m_j}}}.$$

That is, calculating σ_r once takes $O(|G||M|)$ time, which means that the time complexity of Algorithm 3 is $O(k|G||M|)$. In general, k is much less than $|G|$ and $|M|$ in the experiments, so the time complexity of Algorithm 3 is $O(|G||M|)$.

Algorithm 3 Computation of the Parameter σ

Input: Priori knowledge with its length k.
Output: The value of σ.
for $r = 1 : k$ **do**
  Calculate σ_r according to Theorem 1;
end for
$\sigma = \frac{1}{k}\sum_{r=1}^{k}\sigma_r$;
Return σ.

D. Convergence of Our Algorithm

In this section, we discuss the convergence of our algorithm. In Algorithm 1, we adopt a stochastic strategy to obtain a better cognitive result by avoiding the effect of cognition order of attributes. However, it is still necessary to verify whether the cognitive result under the stochastic strategy is consistent with the complete concept of a given clue X_0. Actually, it is equivalent to the following problem: the output (X_p, B_p) of Algorithm 1 can tend to the complete concept $(\mathcal{HL}(X_0), \mathcal{L}(X_0))$ of the given clue X_0 when the running time is long enough. Considering that a concept can uniquely be determined by its intent, this problem is also equivalent to the proposition that the attribute set B_p obtained by the cognition process can tend to the intent $\mathcal{L}(X_0)$ of the complete concept of the given clue when the iteration number is large enough, that is, $B_p \to \mathcal{L}(X_0)$ $(t \to \infty)$.

Theorem 2: Given a clue X_0, let B_p be the set of attributes output by Algorithm 1. Then, $B_p = \mathcal{L}(X_0)$.

Proof: For any $m_k \in B_p$, according to Steps 6 and 21 in Algorithm 1, we have $X_0 \subseteq \mathcal{H}(m_k)$, that is, $m_k \in \mathcal{L}(X_0)$, which leads to $B_p \subseteq \mathcal{L}(X_0)$.

On the other hand, for any $m_k \in \mathcal{L}(X_0)$, by Definition 2, we have $X_0 \subseteq \mathcal{HL}(X_0) \subseteq \mathcal{H}(m_k)$, yielding $\cos(\mathcal{H}(m_k),$

$X_0) \neq 0$. For convenience, suppose that the recognized attributes are $m_1, \ldots, m_l$ in the first l times of cognition. Similarly, $X_0 \subseteq \mathcal{H}(m_i)$ is satisfied for every $m_i \in \{m_1, \ldots, m_l\}$. Thus, $X_0 \subseteq \mathcal{H}(m_i) \cap \mathcal{H}(m_k)$, which leads to $\cos(\mathcal{H}(m_i), \mathcal{H}(m_k)) > 0$. By Definition 9, it follows $P(m_k|m_1, \ldots, m_l, X_0) > 0$. According to the small probability event principle in statistics, the event of selecting m_k in the whole cognition process must happen through multiple repeated experiments. As a result, $m_k \in B_p$ is satisfied. Consequently, $\mathcal{L}(X_0) \subseteq B_p$.

To sum up, we have $B_p = \mathcal{L}(X_0)$. ■

Based on the above discussion, the stochastic selection of attributes in our algorithm not only guarantees that recognizing attributes is an incremental process but also keeps the cognitive result correct. In other words, our method can always obtain an accurate cognitive result as long as the iteration process is sufficient. Therefore, the proposed algorithm is convergent.

V. Experimental Results

In this section, we demonstrate the effectiveness of SI2CCLM and its classification model by conducting comparative experiments. On one hand, we evaluate SI2CCLM by comparing it with other concept-cognitive learning algorithms in terms of cognitive performance. On the other hand, for the classification task, we compare the SI2CCLM-based classification model with 16 machine learning classification methods and four concept-cognitive learning classification methods.

A. Experimental Settings

We totally chose 24 datasets to do experiments. The datasets, Automobile, Harberman, Wdbc, Titanic, and Banana, were taken from the KEEL dataset repository,[1] the datasets Fourclass, Svmguide1, and Ijcnn1 were from the LIBSVM dataset repository,[2] and the rest of datasets were from the UCI dataset repository.[3] The details of these datasets are shown in Table I. In order to ensure the fairness of the experiments and reduce statistical differences, the same training set and testing set were used for all the classification methods, and the experiments were repeated ten times to obtain the average values of evaluation indicators.

Note that all the classification algorithms were trained by MATLAB R2021b and Weka 3.8.6. For fairness, all experiments were implemented on a machine with Intel[4] Xeon[4] Platinum 8259CL CPU @ 2.50 GHz and 256-GB main memory.

B. Evaluation of SI2CCLM

For every dataset, we first need to calculate the parameter σ according to Algorithm 3. For that, the length k of priori knowledge should be determined in advance. Based on the previous discussion on k in Section IV-C, the main objective of considering k is to reduce randomness of the priori knowledge, which is granular concepts $(\mathcal{H}(\vec{g}), \vec{g})$ generated randomly in the experiments. To keep the cognitive effectiveness and control the cost of generating concepts, k was set to 10 for our algorithm in the experiments.

[1] Available at: https://sci2s.ugr.es/keel/datasets.php

[2] Available at: https://www.csie.ntu.edu.tw/~cjlin/libsvmtools/datasets/

[3] Available at: http://archive.ics.uci.edu/

[4] Registered trademark.

TABLE I
Detailed Information of Chosen Datasets

ID	Dataset	#Object	#Feature	#Class
1	Nonverbal	73	22	6
2	Chemical	88	17	2
3	Iris	150	4	3
4	Automobile	159	25	6
5	Qualitative	250	6	2
6	Harberman	306	3	2
7	Wholesale	440	7	2
8	Wdbc	569	30	2
9	Blood	748	4	2
10	Fourclass	862	2	2
11	Raisin	900	7	2
12	Banknote	1372	4	2
13	Anticancer	1850	2	4
14	Steel	1941	33	2
15	Titanic	2201	3	2
16	Chess	3196	36	2
17	Wilt	4839	5	2
18	Banana	5300	2	2
19	Wall	5456	24	4
20	Svmguide1	7089	4	2
21	EEG	14980	14	2
22	Occupancy	20560	7	2
23	Ijcnn1	49990	12	2
24	Secondary	61069	20	4

In Sections III and IV, we applied a stochastic strategy to the recognition of attributes and proved the convergence of SI2CCLM. However, to better balance the accuracy and running time of SI2CCLM in the experiments, it deserves sacrificing accuracy to significantly improve the computational efficiency. In other words, we need to find an appropriate iteration, which is able to achieve the cognition task quickly with a tiny decrease of accuracy. To clarify the above issue, we are going to test the convergence speed of Algorithm 1 on 24 datasets.

We first define the cognitive accuracy as follows:

$$\text{Accu}_{\text{cogn}} = \frac{|B_p \cap B_{\text{full}}|}{|B_p \cup B_{\text{full}}|} \tag{17}$$

where B_p is the set of attributes that Algorithm 1 outputs for a given clue, and B_{full} is the set of attributes obtained by a complete cognition. Obviously, $\text{Accu}_{\text{cogn}} = 1$ indicates that our algorithm achieves a complete cognition; otherwise, our algorithm does not achieve a complete cognition when it is terminated.

Fig. 2 shows the convergence speed of SI2CCLM on 24 datasets. It can be seen from Fig. 2 that the majority of datasets can converge to 1 quickly except Datasets 13, 16, 19, 22, and 24. In other words, the convergence speed of our method is rather satisfactory in most cases within an acceptable iteration number.

Furthermore, to evaluate the cognitive effectiveness of SI2CCLM, we contrasted it with two existing concept-cognitive learning methods: the incremental concept-cognitive learning method in [35], and the upper and lower

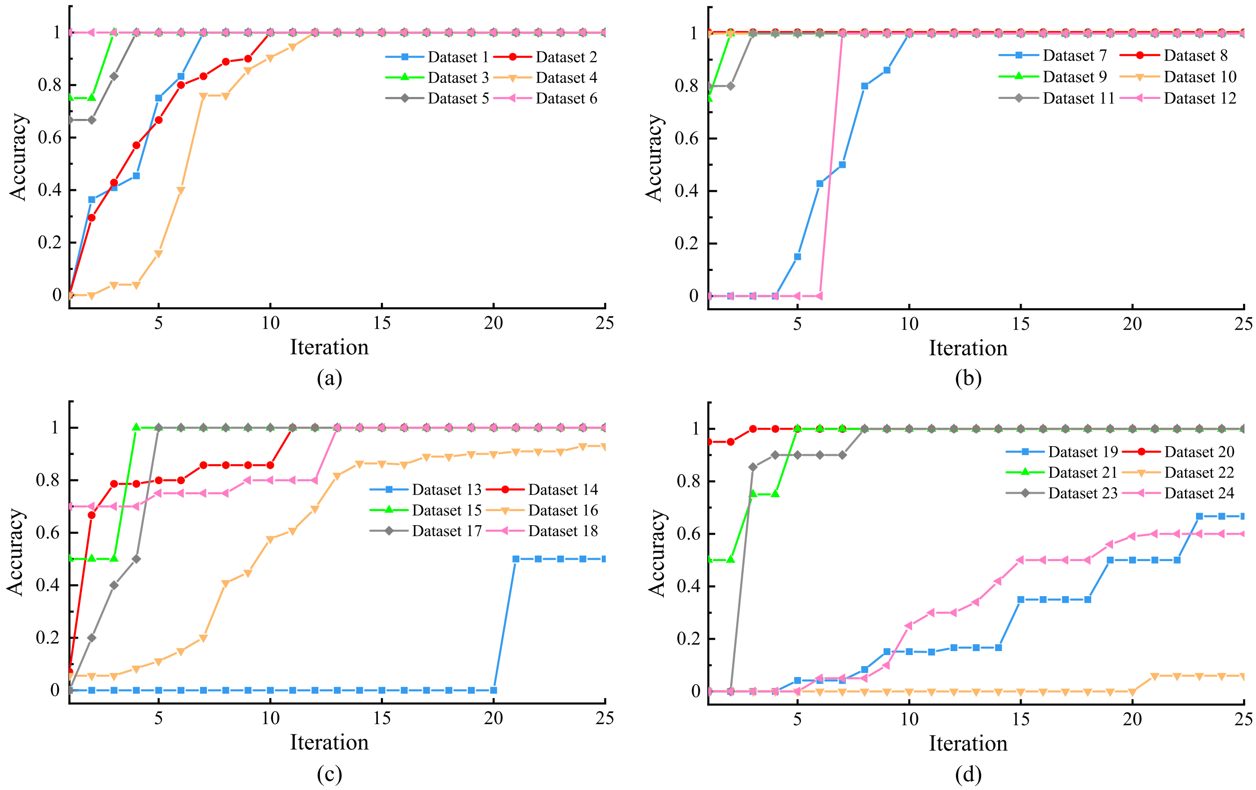

Fig. 2. Convergence speed of SI2CCLM on 24 datasets. (a) Variation in the accuracy of Datasets 1–6. (b) Variation in the accuracy of Datasets 7–12. (c) Variation in the accuracy of Datasets 13–18. (d) Variation in the accuracy of Datasets 19–24.

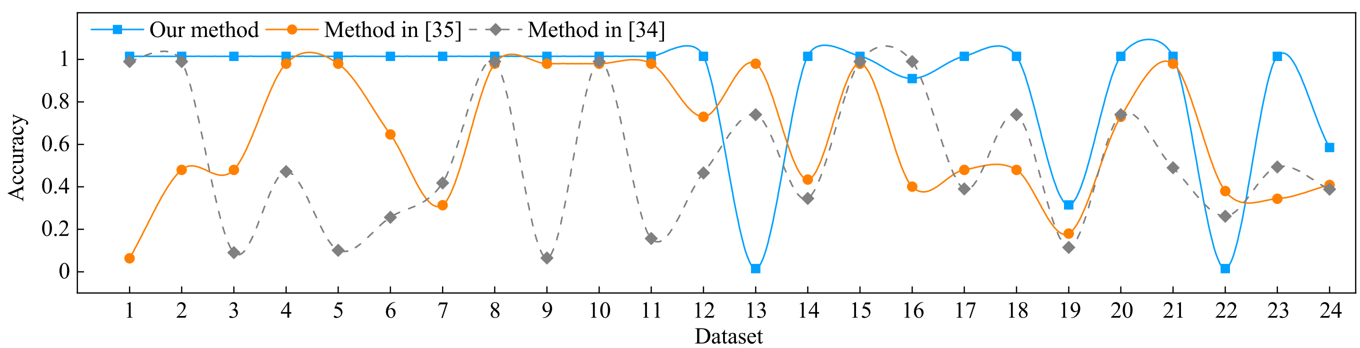

Fig. 3. Comparison of cognitive accuracies of the three different concept-cognitive learning methods.

approximation-based concept-cognitive learning method in [34]. Our experiments took a random object set as a clue from each chosen dataset for concept cognition, and the attributes of the recognized concept were used to calculate cognitive accuracy [i.e., (17)] for achieving the comparison task. It is worth noting that we set the number of iterations as 15 in the experiments to better balance the cognitive accuracy and the running time of SI2CCLM.

For the sake of fairness, we randomly generated a clue with three objects and used the three concept-cognitive learning methods to do concept cognition for this clue. Since the cognition method based on the upper and lower approximations in [34] would get two concepts, we averaged the cognitive accuracies of these two concepts.

Fig. 3 shows the comparison of cognitive accuracies of three concept-cognitive learning methods on 24 datasets. From Fig. 3, our method outperforms the other two concept-cognitive learning methods on 21 datasets and successfully achieves complete cognition on 19 datasets. It is worth noting that the cognitive accuracies of our method become 0 on Datasets 13 and 22. The reason is that the number of iterations for the two datasets was artificially set to 15 as the tradeoff termination condition of our method in the experiments for the purpose of improving the efficiency, that is, the accuracies were not waited to increase until our method has been terminated compulsively. On the other hand, it can be observed from Fig. 2 that the accuracies of our method on Datasets 13 and 22 are indeed 0 in the current iteration number. In fact, it should be noted that although SI2CCLM is convergent, the minority of datasets may converge slowly at the beginning of cognition.

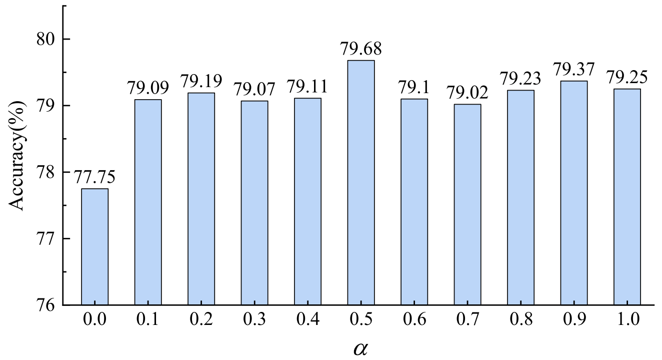

Fig. 4. Influence of the parameter α on the classification accuracies of SI2CCLM-based classification model when $\alpha = 0, 0.1, 0.2, \ldots, 1$.

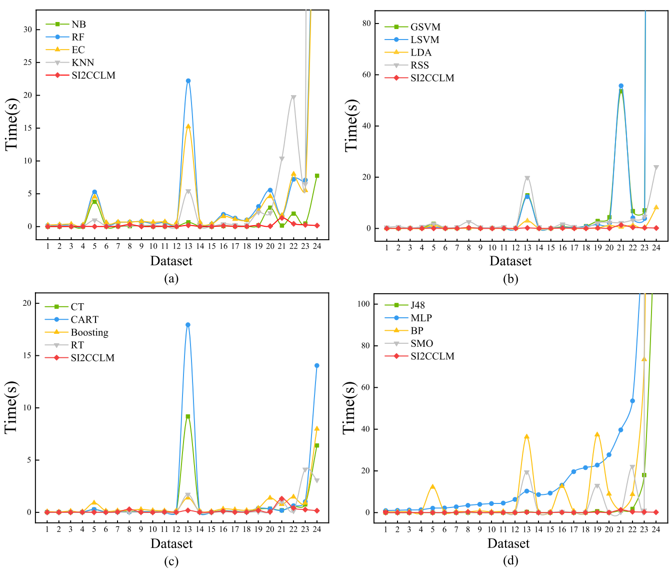

Fig. 5. Comparison of elapsed time of SI2CCLM and machine learning classification methods on 24 datasets. (a) Results of NB, RF, EC, KNN, and SI2CCLM. (b) Results of GSVM, LSVM, LDA, RSS, and SI2CCLM. (c) Results of CT, CART, Boosting, RT, and SI2CCLM. (d) Results of J48, MLP, BP, SMO, and SI2CCLM.

C. Comparison With Other Classification Algorithms

In order to verify the validity of SI2CCLM-based classification model, we compared it with other classification models. For the sake of convenience, we divide the classification methods into two categories: 1) concept-cognitive learning classification methods, namely, CCLM [39], DRCM [41], MACLM [31], and ILMPFTC [26]; 2) machine learning classification models, that is, naive Bayes (NB), random forest (RF), bagged trees in ensemble classifiers (ECs), K-nearest neighbor classifier (KNN), Gaussian kernel function support vector machine (GSVM), linear SVM (LSVM), linear discriminant analysis (LDA), root sum squares (RSS), classification and regression tree (CART), complex tree (CT), boosting, random tree (RT), sequential minimal optimization (SMO), J48, back propagation neural network (BP), and multilayer perceptron (MLP).

For comparison, we divided each dataset into two parts: a training set and a testing set, and used the formula

$$\text{TestRatio} = \frac{|\text{testing set}|}{|\text{training set}| + |\text{testing set}|} \times 100\%$$

TABLE II

AVERAGE CLASSIFICATION ACCURACIES (MEAN ± STANDARD DEVIATION) OF SI2CCLM AND MACHINE LEARNING CLASSIFICATION METHODS

ID	SI2CCLM	NB	RF	EC	KNN	GSVM	LSVM	LDA	RSS	CART	CT	Boosting	RT	SMO	J48	BP	MLP
1	**90.48±0.37**	85.00±6.25●	85.71±5.83●	84.29±4.52●	73.57±6.78●	54.29±13.13●	82.86±4.99●	84.29±4.52●	54.29±13.13●	64.29±13.47●	64.29±13.47●	68.57±10.75●	67.12±0.33●	82.19±0.32●	69.86±0.29●	40.00±14.36●	67.24±0.16●
2	**100.00±0.00**	**100.00±0.00**	**100.00±0.00**	**100.00±0.00**	98.24±2.84●	93.53±4.34●	**100.00±0.00**	98.24±2.84●	81.76±17.42●	**100.00±0.00**	**100.00±0.00**	38.24±6.93●	96.59±0.18●	**100.00±0.00**	**100.00±0.00**	94.71±9.38●	99.63±0.01
3	**98.33±1.76**	97.33±2.63	97.00±2.46	96.67±2.72	90.67±3.44●	96.67±2.72	96.33±3.31	97.67±1.61	94.33±7.71●	96.33±3.31	96.33±3.31	96.33±3.31	92.00±0.21●	92.00±0.31●	92.67±0.20●	96.33±2.92	82.44±0.13●
4	**76.77±2.00**	55.16±3.21●	65.48±4.04●	64.84±6.17●	69.03±8.22●	67.10±3.33●	67.10±3.97●	50.65±3.42●	24.84±8.61●	61.29±6.27●	61.29±6.27●	39.03±6.35●	61.01±0.33●	62.89±0.33●	69.18±0.28●	36.45±8.61●	48.03±0.08●
5	**100.00±0.00**	99.40±0.97	99.80±0.63	**100.00±0.00**	99.80±0.63	**100.00±0.00**	98.20±1.75●	98.00±1.63●	97.00±2.87●	97.80±1.48●	97.80±1.48●	**100.00±0.00**	99.60±0.06	99.60±0.06	99.20±0.09	98.60±1.65	99.47±0.01
6	**82.79±1.56**	78.85±2.94●	81.80±4.40	82.46±3.87	44.92±2.07●	82.30±4.00	78.85±3.41●	79.34±3.39●	80.49±4.19	82.62±3.80	82.62±3.80	79.18±3.63●	76.14±0.43●	74.51±0.50●	76.80±0.42●	**82.79±1.56**	76.37±0.03●
7	**80.68±3.03**	78.41±4.01	78.52±4.23	78.30±3.88	50.45±7.97●	79.09±3.72	78.07±3.86●	78.86±3.18	66.14±4.48●	78.30±3.45	78.30±3.45	78.07±4.08●	77.50±0.41●	78.41±0.46	76.14±0.42●	78.07±3.86●	78.11±0.04
8	**67.88±0.61**	**67.88±0.61**	**67.88±0.61**	**67.88±0.61**	32.12±0.61●	**67.88±0.61**	**67.88±0.61**	**67.88±0.61**	**67.88±2.61**	**67.88±0.61**	**67.88±0.61**	**67.88±0.61**	62.74±0.48●	62.74±0.61●	62.74±0.48●	**67.88±0.61**	62.74±0.04
9	**81.34±1.84**	80.67±1.33	80.87±1.24	81.01±1.31	67.32±11.14●	80.54±1.18	80.67±1.30	80.74±1.65	79.80±1.60	80.40±1.13	80.40±1.13	79.73±1.44	74.73±0.42●	76.07±0.49●	75.94±0.43●	79.60±1.35	75.76±0.03●
10	**90.70±1.13**	85.76±2.18●	**90.70±1.13**	**90.70±1.13**	89.01±1.74	**90.70±1.13**	85.64±2.33●	84.83±2.71●	64.77±2.80●	**90.70±1.13**	**90.70±1.13**	83.95±2.85●	64.39±0.48●	64.39±0.60●	64.39±0.48●	**90.70±1.13**	64.39±0.04●
11	**82.50±1.76**	78.78±1.04●	81.89±1.12	80.56±3.11	77.33±1.50●	82.06±1.83	80.17±1.74●	77.50±1.46●	78.39±0.89●	82.39±1.82	82.39±1.82	80.00±0.87●	82.00±0.36	79.11±0.46●	82.11±0.36	81.00±2.32	82.00±0.03
12	**63.03±1.71**	**63.03±1.71**	61.75±2.89	61.82±2.78	61.06±2.42	**63.03±1.71**	**63.03±1.71**	**63.03±1.71**	58.32±2.57●	**63.03±1.71**	**63.03±1.71**	61.06±2.42	59.77±0.48●	59.77±0.63●	59.77±0.49●	62.37±1.73	59.77±0.03●
13	**91.05±1.62**	83.19±0.57●	90.54±0.44	90.38±0.70	74.11±2.82●	90.16±0.19	86.59±0.34●	**91.05±1.62**	83.19±0.57●	**91.05±1.62**	83.89±0.71●	83.19±0.57●	89.62±0.23●	89.62±0.34●	81.08±0.29●	87.70±1.95●	88.86±0.01●
14	**100.00±0.00**	**100.00±0.00**	**100.00±0.00**	**100.00±0.00**	**100.00±0.00**	**100.00±0.00**	**100.00±0.00**	**100.00±0.00**	64.66±0.70●	**100.00±0.00**	**100.00±0.00**	84.72±1.87●	99.95±0.02	**100.00±0.00**	**100.00±0.00**	**100.00±0.00**	**100.00±0.00**
15	**70.11±1.31**	**70.11±1.31**	**70.11±1.31**	**70.11±1.31**	54.57±20.69●	**70.11±1.31**	**70.11±1.31**	**70.11±1.31**	**70.11±1.31**	**70.11±1.31**	**70.11±1.31**	**70.11±1.31**	67.70±0.47●	67.70±0.57●	67.70±0.47●	**70.11±1.31**	67.70±0.02●
16	95.15±0.55	95.82±0.23	97.51±0.19○	97.53±0.26○	95.67±0.72	95.71±0.25	96.93±0.15○	93.08±0.65●	91.77±0.30●	97.40±0.08○	97.40±0.08○	92.36±0.74●	96.06±0.20	96.65±0.18○	96.50±0.17○	97.51±0.32○	**98.29±0.01**○
17	94.84±0.37	94.96±0.34	**94.98±0.36**	**94.98±0.36**	28.00±1.47	**94.98±0.36**	**94.98±0.36**	**94.98±0.36**	**94.98±0.36**	**94.98±0.36**	**94.98±0.36**	**94.98±0.36**	94.61±0.22	94.61±0.23	94.61±0.23	**94.98±0.36**	94.61±0.01
18	**56.91±0.52**	56.77±0.62	56.77±0.62	56.77±0.62	55.64±0.72	56.77±0.62	56.77±0.62	56.77±0.62	55.92±0.55	56.77±0.62	56.77±0.62	56.12±0.63	55.57±0.49	55.57±0.67	56.02±0.50	56.83±0.71	55.86±0.01
19	90.05±0.27	71.61±1.35●	91.87±0.69	**91.95±0.51**	87.88±0.79●	88.44±0.79●	86.98±0.79●	73.23±1.28●	59.18±3.62●	89.73±0.42●	87.51±0.95●	69.62±0.94●	88.53±0.23●	87.63±0.33●	90.21±0.20	77.33±1.46●	90.38±0.01
20	**59.05±0.86**	**59.05±0.86**	**59.05±0.86**	**59.05±0.86**	57.91±0.77	**59.05±0.86**	**59.05±0.86**	**59.05±0.86**	57.91±0.77	**59.05±0.86**	**59.05±0.86**	**59.05±0.86**	57.50±0.49●	57.50±0.65●	57.50±0.49●	**59.05±0.86**	57.49±0.01●
21	**55.90±0.34**	55.85±0.31	**55.90±0.34**	**55.90±0.34**	**55.90±0.34**	**55.90±0.34**	**55.90±0.34**	**55.90±0.34**	**55.90±0.34**	**55.90±0.34**	**55.90±0.34**	**55.90±0.34**	55.11±0.50	54.09±0.68	55.12±0.50	55.78±0.36	55.11±0.01
22	**99.00±0.08**	98.81±0.11	98.98±0.06	98.98±0.10	96.27±0.12	98.97±0.12	98.98±0.11	98.48±0.13	77.19±0.62●	98.96±0.13	98.96±0.13	98.72±0.25	98.95±0.10	98.98±0.10	98.96±0.10	98.94±0.05	98.95±0.00
23	90.02±0.51	72.22±0.93●	**92.11±0.66**○	91.97±0.49○	87.40±0.97●	88.47±1.28●	87.13±0.80●	73.22±0.99●	59.18±2.67●	90.10±0.76	87.68±0.94●	70.22±0.77●	90.30±0.30	90.31±0.31	90.31±0.30	77.55±1.53●	90.30±0.00
24	**52.01±0.29**	48.72±0.61●	51.75±0.58	51.83±0.57	51.60±0.23	51.53±0.51	51.50±0.23	47.37±0.42●	49.61±0.39●	51.47±0.51	51.26±0.57	49.61±0.39●	51.47±0.36	51.61± 0.38	51.40±0.36	16.07±0.33●	45.81±0.00●
Ave.	**82.02**	78.22	81.29	81.16	70.77	79.47	80.15	78.08	69.48	80.01	79.52	73.19	77.46	78.16	77.84	75.01	76.64

The highest accuracy on each data set is highlighted in boldface, and ●/○ indicates that the performance of SI2CCLM based classification model is significantly better/worse compared to other algorithms by the paired t-test with a confidence level of 95%.

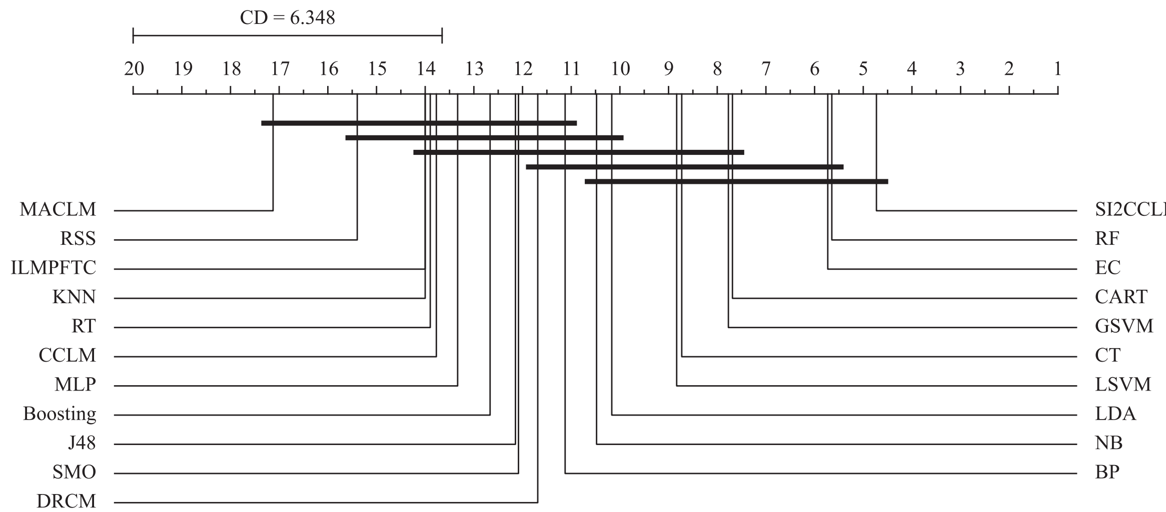

Fig. 6. Diagram of CDs for all the classification methods in the experiments.

to represent the ratio of a testing set to the whole object set. We considered the case of TestRatio $= 0.2$ in the experiments.

Before embarking on the classification comparison, we first discuss the influence of the parameter α on the accuracy of SI2CCLM-based classification model. Fig. 4 shows the effect of the parameter on the average classification accuracies of 24 datasets with $\alpha = 0, 0.1, 0.2, \ldots, 1$. From Fig. 4, the average accuracy of SI2CCLM-based classification model on all datasets reaches the maximum when $\alpha = 0.5$. Hence, we took $\alpha = 0.5$ in the experiments to achieve the comparison task.

To test the performance of SI2CCLM-based classification model, we compared it with 16 machine learning classification algorithms. Table II gives the classification accuracies and standard deviations of SI2CCLM and machine learning classification algorithms on 24 datasets. It can be seen from Table II that the SI2CCLM-based classification model has an average accuracy of 82.02% on 24 datasets, and we can also observe that our model has the highest classification accuracies on all the datasets except Datasets 16, 17, 19, and 23. Furthermore, to further illustrate the advantage of SI2CCLM-based classification model in computational efficiency, we also compare the elapsed time between SI2CCLM and 16 machine learning classification algorithms in Fig. 5. As can be seen from Fig. 5, the elapsed time (red) curves of our model are almost all at the bottom of Fig. 5(a)–(d), which shows that SI2CCLM-based classification model also has advantages in the computational efficiency.

To further demonstrate the effectiveness of SI2CCLM-based classification model, we additionally compared it with four concept-cognitive learning classification algorithms. Table III shows the classification accuracies of SI2CCLM, CCLM, DRCM, MACLM, and ILMPFTC on 24 datasets. It can be seen from Table III that the average classification accuracy of our algorithm is larger than the four compared classification algorithms. Specifically, the classification accuracies of our method are the largest on 18 datasets among all the compared methods. In addition, from Table IV, the average elapsed time of SI2CCLM-based classification model on 24 datasets is less than those of other methods, which demonstrates that the computational efficiency of our algorithm is better than the compared concept-cognitive learning classification algorithms.

TABLE III

AVERAGE CLASSIFICATION ACCURACIES (MEAN ± STANDARD DEVIATION) OF SI2CCLM AND OTHER CLASSIFICATION METHODS BASED ON CONCEPT-COGNITIVE LEARNING

ID	SI2CCLM	CCLM	DRCM	MACLM	ILMPFTC
1	**90.48±0.37**	89.68±5.19	**90.48±3.37**	61.90±12.37●	64.46±0.28●
2	**100.00±0.00**	98.82±2.48	**100.00±0.00**	97.65±4.11	**100.00±0.00**
3	**98.33±1.76**	87.00±4.57●	88.33±5.27●	93.33±8.46●	96.53±0.05●
4	**76.77±2.00**	68.71±5.05●	73.23±5.70●	48.39±2.15●	57.78±0.04●
5	**100.00±0.00**	**100.00±0.00**	99.60±0.84	98.00±2.31●	97.14±0.02●
6	**82.79±1.56**	44.75±4.51●	78.20±4.83●	41.64±6.19●	21.18±0.06●
7	80.68±3.03	35.91±6.70●	80.80±2.86	80.80±2.86	**80.94±0.12**
8	**67.88±0.61**	33.19±1.88●	33.19±1.88●	33.19±1.88●	51.76±0.07●
9	**81.34±1.84**	70.27±4.32●	79.53±2.29●	59.66±6.96●	64.88±0.07●
10	**90.70±1.13**	88.55±2.76●	87.21±1.86●	79.77±1.52●	75.30±0.12●
11	82.50±1.76	76.17±3.23●	77.39±3.22●	67.83±5.82●	**88.67±0.13**○
12	63.03±1.71	59.85±1.62●	59.85±1.62●	57.81±6.23●	**79.07±0.21**○
13	**91.05±1.62**	76.89±1.82●	**91.05±1.62**	10.27±0.90●	89.67±0.06●
14	**100.00±0.00**	**100.00±0.00**	**100.00±0.00**	42.91±1.95●	99.77±0.00
15	**70.11±1.31**	49.18±20.67●	30.41±1.31●	30.41±1.31●	60.38±0.24●
16	95.15±0.55	95.15±0.44	**96.56±0.03**	85.63±0.30●	77.48±0.06●
17	94.84±0.37	33.15±6.28●	**94.95±0.72**	35.15±7.01●	32.83±0.07●
18	56.91±0.52	54.85±2.43●	44.50±0.56●	45.81±3.13●	**61.02±0.01**○
19	**90.05±0.27**	89.28±1.05	90.00±1.03	54.67±1.41●	80.00±0.02●
20	**59.05±0.86**	57.99±0.69	57.99±0.69	57.99±0.69	55.11±0.06●
21	**55.90±0.34**	**55.90±0.34**	**55.90±0.34**	**55.90±0.34**	22.78±0.23●
22	**99.00±0.08**	96.21±0.13●	81.54±0.15●	97.14±0.10●	75.88±0.12●
23	**90.02±0.51**	**90.02±0.51**	10.32±2.14●	85.19±0.95●	86.50±0.04●
24	**52.01±0.29**	51.31±0.29	50.54±0.05●	48.21±2.46●	41.43±0.18●
Ave.	**82.02**	70.95	72.98	61.16	69.19

The highest accuracy on each data set is highlighted in boldface, and ●/○ indicates that the performance of SI2CCLM based classification method is significantly better/worse compared to other algorithms by the paired t-test with a confidence level of 95%.

In order to show the classification performance of our method more intuitively, the average rank and critical difference (CD) of all the classification methods mentioned in the experiments are shown in Fig. 6, where the CD value means the significant difference between the ranking of any two methods at a confidence level of 95%. For the specific calculation process, please refer to [51]. From Fig. 6, it can be seen that our method ranks first among all the compared classification methods, and hence, the performance of SI2CCLM-based classification model is the best.

TABLE IV
COMPARISON OF ELAPSED TIME (S) OF CONCEPT-COGNITIVE LEARNING CLASSIFICATION METHODS ON 24 DATASETS

ID	**SI2CCLM**	**CCLM**	**DRCM**	**MACLM**	**ILMPFTC**
1	0.01	0.02	0.00	0.28	0.01
2	0.02	0.01	0.00	0.26	0.02
3	0.01	0.02	0.00	0.11	0.04
4	0.02	0.03	0.01	0.50	0.07
5	0.01	0.03	0.00	0.87	0.08
6	0.00	0.44	0.02	0.45	0.14
7	0.02	0.93	0.12	2.48	0.14
8	0.15	0.98	0.01	1.61	0.11
9	0.01	0.53	0.01	1.48	0.46
10	0.01	0.63	0.01	2.20	0.14
11	0.03	2.15	0.03	3.00	0.16
12	0.01	8.55	0.03	4.66	6.59
13	0.19	261.00	25.84	75.24	16.61
14	0.02	28.87	0.01	22.59	6.91
15	0.01	42.09	0.02	30.15	10.37
16	0.14	160.63	4.28	287.95	15.99
17	0.01	212.25	0.06	269.84	518.35
18	0.01	202.66	0.03	120.03	4.09
19	0.16	284.26	8.12	117.11	18.10
20	0.02	906.43	0.05	337.16	7.88
21	1.61	13047.71	0.10	6089.61	27.02
22	0.28	12593.52	0.19	22006.90	2705.14
23	0.26	86058.97	0.29	56096.49	20700.31
24	0.16	435887.74	50.10	100987.01	5244.01
Ave.	**0.13**	22904.18	3.72	7769.08	1220.11

VI. CONCLUSION

In this article, we have proposed an incremental incomplete concept-cognitive learning method by using a stochastic strategy. The stochastic cognition of concepts was realized by calculating the probabilities of attributes being selected in priority. In addition, we have constructed a classification model based on the SI2CCLM whose convergence has been proved as well. Compared with the existing concept-cognitive learning algorithms and machine learning classification algorithms, we have demonstrated the validity of our method from the perspectives of cognitive accuracy, classification accuracy, and running time.

The current work can provide a reference for the further study of concept-cognitive learning from a stochastic viewpoint. Such kind of thinking can effectively avoid the effect of the cognition order of attributes on the cognitive results, and hence, it is able to achieve better concept cognition and classification performance.

However, people's cognition of concepts is very complex in reality, that is, although this article has made a preliminary discussion on the stochastic cognition of concepts by using a stochastic strategy, it is still not enough because cognition in some cases cannot be realized by an individual cognitive subject. For example, different cognitive subjects may have different cognitive results for the same clue, and a global consistent cognitive result often requires the fusion of different cognitive results obtained by different individual cognitive subjects. Nevertheless, the stochastic incremental incomplete CCLM depends on priori knowledge, which is granular concepts generated randomly in the experiments. However, it is not sure whether the granular concepts are qualified representatives since granular concepts are only a small portion of concepts. In addition, to explore the practical application of our method, it is still necessary and important to find more specific scenes to illustrate our method. The above problems deserve to be studied in our future research.

REFERENCES

[1] Y. X. Wang, "On cognitive computing," *Int. J. Softw. Sci. Comput. Intell.*, vol. 1, no. 3, pp. 1–15, Jul. 2009.

[2] D. S. Modha, R. Ananthanarayanan, S. K. Esser, A. Ndirango, A. J. Sherbondy, and R. Singh, "Cognitive computing," *Commun. ACM*, vol. 54, no. 8, pp. 62–71, Aug. 2011.

[3] J. Feldman, "Minimization of Boolean complexity in human concept learning," *Nature*, vol. 407, no. 6804, pp. 630–633, Oct. 2000.

[4] B. M. Lake, R. Salakhutdinov, and J. B. Tenenbaum, "Human-level concept learning through probabilistic program induction," *Science*, vol. 350, no. 6266, pp. 1332–1338, Dec. 2015.

[5] Y. X. Wang and Y. Wang, "Cognitive informatics models of the brain," *IEEE Trans. Syst., Man, Cybern., C*, vol. 36, no. 2, pp. 203–207, Mar. 2006.

[6] Y. X. Wang, "On concept algebra: A denotational mathematical structure for knowledge and software modeling," *Int. J. Cogn. Informat. Nat. Intell.*, vol. 2, no. 2, pp. 1–19, Apr. 2008.

[7] W.-Z. Wu, Y. Leung, and J.-S. Mi, "Granular computing and knowledge reduction in formal contexts," *IEEE Trans. Knowl. Data Eng.*, vol. 21, no. 10, pp. 1461–1474, Oct. 2009.

[8] J. T. Yao, A. V. Vasilakos, and W. Pedrycz, "Granular computing: Perspectives and challenges," *IEEE Trans. Cybern.*, vol. 43, no. 6, pp. 1977–1989, Dec. 2013.

[9] L. Wei and Q. Wan, "Granular transformation and irreducible element judgment theory based on pictorial diagrams," *IEEE Trans. Cybern.*, vol. 46, no. 2, pp. 380–387, Feb. 2016.

[10] S. Xia, S. Zheng, G. Wang, X. Gao, and B. Wang, "Granular ball sampling for noisy label classification or imbalanced classification," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 34, no. 4, pp. 2144–2155, Apr. 2023.

[11] S. Xia, X. Dai, G. Wang, X. Gao, and E. Giem, "An efficient and adaptive granular-ball generation method in classification problem," *IEEE Trans. Neural Netw. Learn. Syst.*, early access, Oct. 5, 2022, doi: 10.1109/TNNLS.2022.3203381.

[12] C. Aswani Kumar, M. S. Ishwarya, and C. K. Loo, "Formal concept analysis approach to cognitive functionalities of bidirectional associative memory," *Biologically Inspired Cognit. Architectures*, vol. 12, pp. 20–33, Apr. 2015.

[13] W. Xu and W. Li, "Granular computing approach to two-way learning based on formal concept analysis in fuzzy datasets," *IEEE Trans. Cybern.*, vol. 46, no. 2, pp. 366–379, Feb. 2016.

[14] F. Hao, G. Min, Z. Pei, D.-S. Park, and L. T. Yang, "K-clique community detection in social networks based on formal concept analysis," *IEEE Syst. J.*, vol. 11, no. 1, pp. 250–259, Jun. 2017.

[15] T. Zhang, H.-H. Li, M.-Q. Liu, and M. Rong, "Incremental concept-cognitive learning based on attribute topology," *Int. J. Approx. Reasoning*, vol. 118, pp. 173–189, Mar. 2020.

[16] Z. Pawlak, "Rough set theory and its applications to data analysis," *Cybern. Syst.*, vol. 29, no. 7, pp. 661–688, Oct. 1998.

[17] C. Wang, Q. Hu, X. Wang, D. Chen, Y. Qian, and Z. Dong, "Feature selection based on neighborhood discrimination index," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 29, no. 7, pp. 2986–2999, Jul. 2018.

[18] W. Wei and J. Liang, "Information fusion in rough set theory: An overview," *Inf. Fusion*, vol. 48, pp. 107–118, Aug. 2019.

[19] W. T. Li, H. X. Zhou, W. H. Xu, X. Z. Wang, and W. Pedrycz, "Interval dominance-based feature selection for interval-valued ordered data," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 34, no. 10, pp. 6898–6912, Oct. 2023.

[20] Y. Yao, "Interpreting concept learning in cognitive informatics and granular computing," *IEEE Trans. Syst., Man, Cybern., B*, vol. 39, no. 4, pp. 855–866, Aug. 2009.

[21] G. Qiu, J. Ma, H. Yang, and W. Zhang, "A mathematical model for concept granular computing systems," *Sci. China Inf. Sci.*, vol. 53, no. 7, pp. 1397–1408, Jul. 2010.

[22] Y. Zhao, J. Li, W. Liu, and W. Xu, "Cognitive concept learning from incomplete information," *Int. J. Mach. Learn. Cybern.*, vol. 8, no. 1, pp. 159–170, Feb. 2017.

[23] B. Fan, E. C. C. Tsang, W. Xu, D. Chen, and W. Li, "Attribute-oriented cognitive concept learning strategy: A multi-level method," *Int. J. Mach. Learn. Cybern.*, vol. 10, no. 9, pp. 2421–2437, Sep. 2019.

[24] C. Huang, J. Li, C. Mei, and W.-Z. Wu, "Three-way concept learning based on cognitive operators: An information fusion viewpoint," *Int. J. Approx. Reasoning*, vol. 83, pp. 218–242, Apr. 2017.

[25] J. Li, C. Huang, J. Qi, Y. Qian, and W. Liu, "Three-way cognitive concept learning via multi-granularity," *Inf. Sci.*, vol. 378, pp. 244–263, Feb. 2017.

[26] K. Yuan, W. Xu, W. Li, and W. Ding, "An incremental learning mechanism for object classification based on progressive fuzzy three-way concept," *Inf. Sci.*, vol. 584, pp. 127–147, Jan. 2022.

[27] Y. Mi, Y. Shi, J. Li, W. Liu, and M. Yan, "Fuzzy-based concept learning method: Exploiting data with fuzzy conceptual clustering," *IEEE Trans. Cybern.*, vol. 52, no. 1, pp. 582–593, Jan. 2022.

[28] J. Niu, D. Chen, J. Li, and H. Wang, "Fuzzy rule-based classification method for incremental rule learning," *IEEE Trans. Fuzzy Syst.*, vol. 30, no. 9, pp. 3748–3761, Sep. 2022.

[29] W. H. Xu, D. D. Guo, Y. H. Qian, and W. P. Ding, "Two-way concept-cognitive learning method: A fuzzy-based progressive learning," *IEEE Trans. Fuzzy Syst.*, vol. 31, no. 6, pp. 1885–1899, Jun. 2023.

[30] T. Zhang, M. Rong, H. Shan, and M. Liu, "Causal asymmetry analysis in the view of concept-cognitive learning by incremental concept tree," *Cognit. Comput.*, vol. 13, no. 5, pp. 1274–1286, Sep. 2021.

[31] W. Xu and Y. Chen, "Multi-attention concept-cognitive learning model: A perspective from conceptual clustering," *Knowl.-Based Syst.*, vol. 252, Sep. 2022, Art. no. 109472.

[32] W. X. Zhang and W. H. Xu, "Cognitive model based on granular computing," *Chin. J. Eng. Math.*, vol. 24, pp. 957–971, Jun. 2007.

[33] W. Xu, J. Pang, and S. Luo, "A novel cognitive system model and approach to transformation of information granules," *Int. J. Approx. Reasoning*, vol. 55, no. 3, pp. 853–866, Mar. 2014.

[34] J. Li, C. Mei, W. Xu, and Y. Qian, "Concept learning via granular computing: A cognitive viewpoint," *Inf. Sci.*, vol. 298, pp. 447–467, Mar. 2015.

[35] J. H. Li, Y. L. Mi, and W. Q. Liu, "Incremental cognition of concepts: Theories and methods," *Chin. J. Comput.*, vol. 42, no. 10, pp. 2233–2250, Oct. 2019.

[36] Q. Hu, J. Mi, and D. Chen, "Granular computing based machine learning in the era of big data," *Inf. Sci.*, vol. 378, pp. 242–243, Feb. 2017.

[37] Y. Mi, W. Liu, Y. Shi, and J. Li, "Semi-supervised concept learning by concept-cognitive learning and concept space," *IEEE Trans. Knowl. Data Eng.*, vol. 34, no. 5, pp. 2429–2442, May 2022.

[38] Y. L. Mi, J. H. Li, W. Q. Liu, and J. Lin, "Research on granular concept cognitive learning system under MapReduce framework," *Acta Electronica Sinica*, vol. 46, no. 2, pp. 289–297, Feb. 2018.

[39] Y. Shi, Y. Mi, J. Li, and W. Liu, "Concept-cognitive learning model for incremental concept learning," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 51, no. 2, pp. 809–821, Feb. 2021.

[40] Y. Shi, Y. Mi, J. Li, and W. Liu, "Concurrent concept-cognitive learning model for classification," *Inf. Sci.*, vol. 496, pp. 65–81, Sep. 2019.

[41] J. Niu, D. Chen, J. Li, and H. Wang, "A dynamic rule-based classification model via granular computing," *Inf. Sci.*, vol. 584, pp. 325–341, Jan. 2022.

[42] R. W. Scholz, *Cognitive Strategies in Stochastic Thinking*. Dodrecht, Holland: D. Reidel, 1987.

[43] M. Bar-Hillel and W. A. Wagenaar, "The perception of randomness," *Adv. Appl. Math.*, vol. 12, no. 4, pp. 428–454, Dec. 1991.

[44] R. S. Nickerson, "The production and perception of randomness," *Psychol. Rev.*, vol. 109, no. 2, pp. 330–357, 2002.

[45] T. Meiser, "Analyzing stochastic dependence of cognitive processes in multidimensional source recognition," *Experim. Psychol.*, vol. 61, no. 5, pp. 402–415, May 2014.

[46] P. A. Warren, U. Gostoli, G. D. Farmer, W. El-Deredy, and U. Hahn, "A re-examination of 'bias' in human randomness perception," *J. Exp. Psychol. Hum.*, vol. 44, no. 5, pp. 663–680, May 2018.

[47] R. Wille, "Restructuring lattice theory: An approach based on hierarchies of concepts," in *Ordered Sets*, I. Rival, Ed. Dordrecht, The Netherlands: Springer, Sep. 1982, pp. 445–470.

[48] Z. H. Zhou, *Machine Learning*. Beijing, China: Tsinghua Univ. Press, Jan. 2016.

[49] P. N. Tan, M. Steinbach, and V. Kumar, *Introduction to Data Mining*. Boston, MA, USA: Addison-Wesley, 2006.

[50] J. Chen, I. Y. Liao, B. Belaton, and M. Zaman, "Division-based large point set registration using coherent point drift (CPD) with automatic parameter tuning1," *J. Intell. Fuzzy Syst.*, vol. 28, no. 5, pp. 2297–2308, Feb. 2015.

[51] J. Demšar, "Statistical comparisons of classifiers over multiple data sets," *J. Mach. Learn. Res.*, vol. 7, pp. 1–30, Jan. 2006.

Zhiming Liu received the B.Sc. degree in mathematics from Shanxi University, Taiyuan, China, in 2018. She is currently pursuing the Ph.D. degree with the Kunming University of Science and Technology, Kunming, China.

Her research interests include machine learning, concept-cognitive learning, and granular computing.

Jinhai Li received the M.Sc. degree in science from Guangxi University, Nanning, China, in 2009, and the Ph.D. degree in science from Xi'an Jiaotong University, Xi'an, China, in 2012.

He is currently a Professor with the Kunming University of Science and Technology, Kunming, China. In recent years, he has published more than 40 articles in IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, IEEE TRANSACTIONS ON CYBERNETICS, IEEE TRANSACTIONS ON FUZZY SYSTEMS, IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS: SYSTEM, and *Pattern Recognition*. His research interests include big data, cognitive computing, granular computing, and formal concept analysis.

Dr. Li is also an Area Editor of *International Journal of Approximate Reasoning* and an Associate Editor of *International Journal of Machine Learning and Cybernetics*.

Xiao Zhang received the B.Sc. degree in science from Henan Normal University, Xinxiang, China, in 2008, the M.Sc. degree in science from North China Electric Power University, Beijing, China, in 2011, and the Ph.D. degree in science from Xi'an Jiaotong University, Xi'an, China, in 2014.

She is currently an Associate Professor with the School of Science, Xi'an University of Technology, Xi'an. Her current research interests include rough sets, granular computing, data mining, and machine learning.

Xi-Zhao Wang (Fellow, IEEE) received the B.Sc. degree in mathematics from Hebei University, Shijiazhuang, China, in 1983, the M.Sc. degree in computer science from Shanghai Jiao Tong University, Shanghai, China, in 1987, and the Ph.D. degree in computer science from the Harbin Institute of Technology, Harbin, China, in 1998.

He is currently a Professor with the School of Computer and Software, Shenzhen University, Shenzhen, China. His main research interests include machine learning and uncertainty information processing, including inductive learning with fuzzy representation, approximate reasoning and expert systems, neural networks and their sensitivity analysis, statistical learning theory, fuzzy measures and fuzzy integrals, random weight networks, and machine learning theories and methodologies in the big-data environment.

Dr. Wang is a fellow of Chinese Association for Artificial Intelligence (CAAI). He serves on the Advisory Board of IEEE TRANSACTIONS ON FUZZY SYSTEMS and is a member of a number of editorial boards of other international journals. He is also the Editor-in-Chief of *International Journal of Machine Learning and Cybernetics*, the Deputy Chair of the CAAI Machine Learning Committee and the Knowledge Engineering Committee, the Overseas High-Level (peacock B class) Talent of Shenzhen City, the Prizewinner of the First-Class Award of the Natural Science Advances of Hebei Province, and a Model Teacher of China.