Knowledge-Based Systems 290 (2024) 111566

Contents lists available at ScienceDirect

Knowledge-Based Systems

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

Correlation concept-cognitive learning model for multi-label classification

Jiaming Wu [a], Eric C.C. Tsang [a,*], Weihua Xu [b], Chengling Zhang [a], Lanzhen Yang [a,c]

[a] School of Computer Science and Engineering, Macau University of Science and Technology, Taipa, Macau, China
[b] College of Artificial Intelligence, Southwest University, Chongqing, 400715, China
[c] College of Mathematics and Information Sciences, Hebei University, Baoding, 071002, China

ARTICLE INFO

Keywords:
Concept-cognitive learning
Granular computing
Formal concept analysis
Multi-label learning
Label correlations

ABSTRACT

As a cognitive process, concept-cognitive learning (CCL) emphasizes the structured expression of data through systematic cognition and understanding, to obtain valuable information in the data. Although concept-cognitive learning has achieved good results in single-label classification tasks, it has not yet been applied to multi-label learning. The difficulty is that the existing concept-cognitive learning fails to effectively associate and utilize the structural relationships in the feature concepts and the multi-label concepts. In addition, in the face of complex multi-label datasets, it is difficult to obtain effective concepts for classification tasks. To solve these problems, this paper proposes a correlation concept-cognitive learning method and applies it to the multi-label classification (MLC) task. Moreover, the relationship between the feature concepts and the multi-label concepts is established by extent. On this basis, we comprehensively consider the extent relevancy and intent relevancy, and learn correlation concepts. In order to improve the classification precision, we construct correlation concept spaces to obtain the representation of effective concepts. Finally, we conducted experimental evaluations on ten datasets to illustrate the effectivity and advantages of the proposed approach.

1. Introduction

Concept-cognitive learning refers to the process of acquiring abstract concepts and classifying concepts by simulating human cognitive processes and psychological mechanisms [1]. It is also a data science that focuses on obtaining valuable knowledge from data. It is frequently utilized in machine learning [2,3], cognitive psychology [4,5], rule extraction [6,7] and other fields.

In the 1980s, Wille proposed formal concept analysis (FCA) [8], which establishes the structure and relationship of concepts through a formal method. FCA is based on lattice theory and set theory. A classical concept usually consists of two parts, namely extent (object set) and intent (feature set), which can be mutually determined [9]. As the application scope of FCA gradually expands, researchers have successively proposed fuzzy concepts [10–12], three-way concepts [13], multi-scale concepts [14], etc. Granular computing is a way of intelligent information processing, which can deal with problems at multiple granularity levels [15–19]. Yao [20] believes that the formation and learning of concepts is the central issue of granular computing and cognitive informatics, and constructs the framework of concept learning from the philosophical, methodological and application levels, and effectively combines granular computing with concept learning. In addition, Wu et al. [21] introduced the idea of granular computing and knowledge reduction to divide the data into different granularities and simplify the concept learning process. Last several years, with the better understanding of CCL, more and more scholars start to do research in this field. Concretely, Shi [22,23] and Mi [24–26] proposed a variety of concept-cognitive learning models and used them to deal with classification tasks. Considering the limitations of individual cognition and the incompleteness of cognitive environment, Yuan et al. [27] constructed a three-way fuzzy progressive idea for classifying objects. When viewed via the lens of concept clustering, Xu et al. [28] put forward a multi-attention concept-cognitive learning model through graph attention mechanism. Zhang et al. [29] weighted concepts according to the importance of different attributes, and removing duplicate information with a progressive weighted fuzzy concept. Although the aforementioned techniques have achieved good results in classification tasks, they all have a common problem that they can only handle single-label tasks. In the data era, this greatly limits the development of concept-cognitive learning.

Different from traditional single-label classification, the multi-label learning (MLL) task is to predict multiple labels for each sample. In MLL, each sample is linked to multiple labels, which may or may not be related [30]. With the development of the data era, various fields have employed multi-label learning extensively, and many multi-label

* Corresponding author.
E-mail addresses: wjmfuzzy@163.com (J. Wu), cctsang@must.edu.mo (E.C.C. Tsang), chxuwh@gmail.com (W. Xu), chengling_zhang@126.com (C. Zhang), lzyang@hbu.edu.cn (L. Yang).

https://doi.org/10.1016/j.knosys.2024.111566
Received 17 July 2023; Received in revised form 16 January 2024; Accepted 23 February 2024
Available online 24 February 2024
0950-7051/

learning algorithms have emerged [31–36]. Two types of multi-label learning exist: problem transformation and algorithm adaptation. The problem transformation method converts the multi-label learning problem into multiple independent single-label problems, and then uses single-label learning algorithms to deal with them, such as BR [37], LIFT [38], etc. However, these methods often ignore the correlation between labels, which may lead to inaccurate prediction results. The other type is algorithm adaptation methods which usually modified or extended by single-label classification algorithms, which can be applied directly to process multi-label datasets, such as ML-KNN [39], CC [40], RAkEL [41], etc. With deep understanding of this research, the effectiveness of multi-label classification can be enhanced by measuring and using the relevance between labels and features as well as the relevance between labels [42]. More and more scholars begin to study the problem of label correlation [43–45]. LLSF-DL [46] uses sparse superposition to exploit high-order label correlation. MSWL [47] uses a manifold regularized sparse model to mine the correlation between labels and feature structures. GLMAM [48] evaluates label correlation globally and locally, and introduces an attention mechanism to modify the original label space by encoding instance and label information. GLFS [49] uses local correlation to improve generalization performance, and uses label group and instance group correlation to promote model training. 2SML [50] uses feature manifold and label manifold to share a weight matrix by using the prior knowledge of correlation. Although correlation has made some progress in multi-label learning, there are still some problems and challenges.

Based on the above analysis, the challenges are summarized as follows: (1) The limitation of traditional CCL lies in its failure to consider the structural relationships within the label space, making it challenging for conventional methods to directly handle multi-label data. (2) Most of the existing correlation studies focus on the degree of correlation between labels, while ignoring the influence of objects and features on the correlation between labels. (3) It is difficult to obtain effective concepts for classification tasks in the complex multi-label datasets.

In view of the above challenges, this paper combines concept-cognitive learning theory with multi-label learning to construct correlation concept that considers the correlation between objects, features and labels and predicts multi-label datasets. Fig. 1 displays the primary framework of the proposed approach. In addition, the following are the innovation and the primary contributions for the research:

(1) By using extent information as a connecting bridge, we establish a close association between multi-label concepts and feature concepts, thereby constructing a novel multi-label feature concept space. This space effectively characterizes the intricate interactions and relationships between multi-label concepts and feature concepts.

(2) We introduce an innovative concept, termed "the correlation concept". It concurrently explores both extent relevancy and intent relevancy aiming to comprehensively depict the intricate structures and interconnections among objects, labels and key features. It provides a more comprehensive methodology for high-order correlation analysis.

(3) We employ both positive and negative perspectives of correlation concepts for label prediction, comprehensively mining knowledge information, and thereby minimizing the impact of cognitive bias. Through a series of experimental validations, the results clearly demonstrate the significant effectiveness and superiority of this approach over other approaches.

The rest of this essay is structured as follows. Section 2 reviews the basic knowledge of CCL and MLL. In Section 3, we addresses the cognitive learning process of correlation concept space and label prediction algorithm in detail. The Section 4 conducts numerical experiments. Finally, the conclusion and future work are discussed in Section 5.

2. Preliminaries

Concept-cognitive learning and multi-label learning are briefly reviewed in this part. More detailed information could be found in [1, 13,30].

2.1. Concept-cognitive learning

Definition 1. Let (U, A, R) be a regular formal context, an object set is represented by U, and a feature set by A. The binary relationship $R : U \times A \to \{0,1\}$ between U and A, that is, $xRa = 1$ represents object x with feature a. The power sets of U and A are represented by 2^U and 2^A, respectively, then $\mathcal{L}^c : 2^U \to 2^A$ and $\mathcal{H}^c : 2^A \to 2^U$ are two set-valued mappings. In addition, for $X \subseteq U, B \subseteq A$, the positive operators $\mathcal{L}^c$ and $\mathcal{H}^c$ are defined:

$$\mathcal{L}^c(X) = \{a \in A | \forall x \in X, xRa = 1\}, \tag{1}$$

$$\mathcal{H}^c(B) = \{x \in U | \forall a \in B, xRa = 1\}. \tag{2}$$

If the binary group (X, B) satisfies $\mathcal{L}^c(X) = B$ and $\mathcal{H}^c(B) = X$, then (X, B) is referred to as a positive feature concept. Where B is the intent of the concept, that is, the maximal set of the features that all the objects in X have in common. X is the extent of the positive concept, that is, the maximal set of the objects shared by all the features in B.

Similarly, $xRa = 0$ represents object x without feature a. And $\bar{\mathcal{L}}^c : 2^U \to 2^A$ and $\bar{\mathcal{H}}^c : 2^A \to 2^U$ are two set-valued mappings. For $\overline{X} \subseteq U, \overline{B} \subseteq A$, the negative operators $\bar{\mathcal{L}}^c$ and $\bar{\mathcal{H}}^c$ are defined as follows:

$$\bar{\mathcal{L}}^c\left(\overline{X}\right) = \left\{a \in A | \forall x \in \overline{X}, xRa = 0\right\}, \tag{3}$$

$$\bar{\mathcal{H}}^c\left(\overline{B}\right) = \left\{x \in U | \forall a \in \overline{B}, xRa = 0\right\}. \tag{4}$$

If the binary group $\left(\overline{X}, \overline{B}\right)$ satisfies $\bar{\mathcal{L}}^c\left(\overline{X}\right) = \overline{B}$ and $\bar{\mathcal{H}}^c\left(\overline{B}\right) = \overline{X}$, then $\left(\overline{X}, \overline{B}\right)$ is referred to as a negative feature concept. The $\overline{B}$ and $\overline{X}$ are called the intent and extent of this negative concept, respectively. The negative concept could induce the information that object and feature do not have in common.

Proposition 1. *For any $X_1, X_2 \subseteq U$ and $B \subseteq A$ hold the following properties:*

$$X_1 \subseteq X_2 \Rightarrow \mathcal{L}^c\left(X_2\right) \subseteq \mathcal{L}^c\left(X_1\right),$$
$$\mathcal{L}^c\left(X_1 \cup X_2\right) \supseteq \mathcal{L}^c\left(X_1\right) \cap \mathcal{L}^c\left(X_2\right),$$
$$\mathcal{H}^c(B) = \{x \in U | B \subseteq \mathcal{L}^c(\{x\})\}.$$

Definition 2. In regular formal context (U, A, R), for any $x \in U$ and $a \in A$, $(\mathcal{H}^c\mathcal{L}^c(x), \mathcal{L}^c(x))$ and $(\mathcal{H}^c(a), \mathcal{L}^c\mathcal{H}^c(a))$ are called positive granular concepts under the positive operators $\mathcal{L}^c$ and $\mathcal{H}^c$, $\left(\bar{\mathcal{H}}^c\bar{\mathcal{L}}^c(x), \bar{\mathcal{L}}^c(x)\right)$ and $\left(\bar{\mathcal{H}}^c(a), \bar{\mathcal{L}}^c\bar{\mathcal{H}}^c(a)\right)$ are called negative granular concepts under the negative operators $\bar{\mathcal{L}}^c$ and $\bar{\mathcal{H}}^c$. Moreover, we denote the positive concept space by $\mathcal{G}_{\mathcal{L}^c\mathcal{H}^c}$ and the negative concept space by $\mathcal{G}_{\bar{\mathcal{L}}^c\bar{\mathcal{H}}^c}$, which are the sets of all positive granular concepts and negative granular concepts, respectively, that is:

$$\mathcal{G}_{\mathcal{L}^c\mathcal{H}^c} = \{(\mathcal{H}^c\mathcal{L}^c(x), \mathcal{L}^c(x)) | x \in U\} \cup \{(\mathcal{H}^c(a), \mathcal{L}^c\mathcal{H}^c(a)) | a \in A\}, \tag{5}$$

$$\mathcal{G}_{\bar{\mathcal{L}}^c\bar{\mathcal{H}}^c} = \left\{\left(\bar{\mathcal{H}}^c\bar{\mathcal{L}}^c(x), \bar{\mathcal{L}}^c(x)\right) | x \in U\right\} \cup \left\{\left(\bar{\mathcal{H}}^c(a), \bar{\mathcal{L}}^c\bar{\mathcal{H}}^c(a)\right) | a \in A\right\}. \tag{6}$$

Example 1. Table 1 is a regular formal context with 10 objects and 7 features, where $U = \{x_1, x_2, \ldots, x_{10}\}$ represents 10 kinds of animals, representing chicken, whale, cat, centipede, butterfly, tiger, eagle, scorpion, leopard and ostrich. $A = \{a_1, a_2, \ldots, a_7\}$ represents 7 features, respectively, constant temperature, lung breathing, feathers, fur, spawning, vertebral bone and dangerous.

For this regular formal context, the positive concept space $\mathcal{G}_{\mathcal{L}^c\mathcal{H}^c}$ and the negative concept space $\mathcal{G}_{\bar{\mathcal{L}}^c\bar{\mathcal{H}}^c}$ are shown as follows:

$$\mathcal{G}_{\mathcal{L}^c\mathcal{H}^c} = \left\{\begin{array}{c} \left(\{x_1, x_7, x_{10}\}, \{a_1, a_2, a_3, a_5, a_6\}\right), \left(\{x_2, x_6, x_7, x_9\},\right. \\ \left.\{a_1, a_2, a_6, a_7\}\right), \left(\{x_3, x_6, x_9\}, \{a_1, a_2, a_4, a_6\}\right), \\ \left(\{x_4, x_7, x_8\}, \{a_5, a_7\}\right), \left(\{x_1, x_4, x_5, x_7, x_8, x_{10}\}, \{a_5\}\right), \\ \left(\{x_6, x_9\}, \{a_1, a_2, a_4, a_6, a_7\}\right), \\ \left(\{x_7\}, \{a_1, a_2, a_3, a_5, a_6, a_7\}\right), \left(\{x_1, x_2, x_3, x_6, x_7, x_9, x_{10}\},\right. \\ \left.\{a_1, a_2, a_6\}\right), \\ \left(\{x_2, x_4, x_6, x_7, x_8, x_9\}, \{a_7\}\right) \end{array}\right\},$$

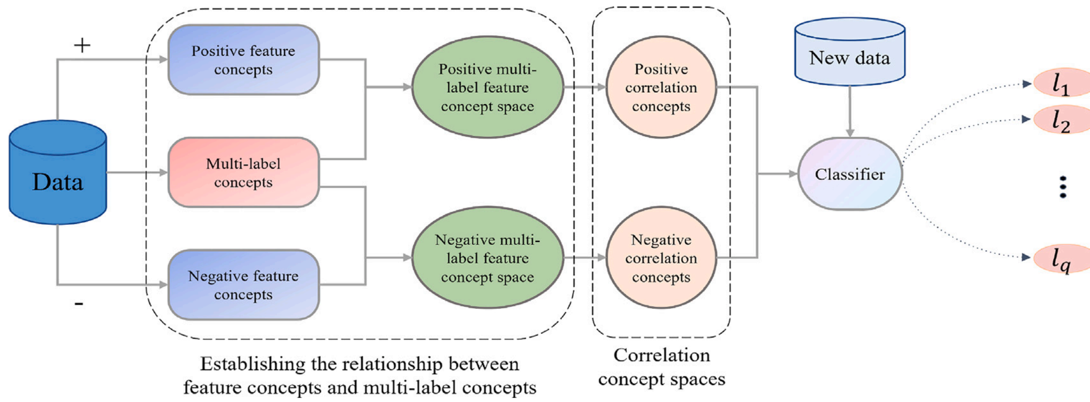


Fig. 1. The main framework of the proposed method.

Table 1
A regular formal context.

U	a_1	a_2	a_3	a_4	a_5	a_6	a_7
x_1	1	1	1	0	1	1	0
x_2	1	1	0	0	0	1	1
x_3	1	1	0	1	0	1	0
x_4	0	0	0	0	1	0	1
x_5	0	0	0	0	1	0	0
x_6	1	1	0	1	0	1	1
x_7	1	1	1	0	1	1	1
x_8	0	0	0	0	1	0	1
x_9	1	1	0	1	0	1	1
x_{10}	1	1	1	0	1	1	0

$$\mathcal{G}_{\bar{\mathcal{L}}^c\bar{\mathcal{H}}^c} = \left\{ \begin{array}{c} \left(\left\{x_1, x_5, x_{10}\right\}, \left\{a_4, a_7\right\}\right), \left(\left\{x_2\right\}, \left\{a_3, a_4, a_5\right\}\right), \\ \left(\left\{x_3\right\}, \left\{a_3, a_5, a_7\right\}\right), \\ \left(\left\{x_4, x_5, x_8\right\}, \left\{a_1, a_2, a_3, a_4, a_6\right\}\right), \left(\left\{x_5\right\}, \left\{a_1, a_2, a_3, a_4, a_6, a_7\right\}\right), \\ \left(\left\{x_2, x_3, x_6, x_9\right\}, \left\{a_3, a_5\right\}\right), \\ \left(\left\{x_1, x_2, x_4, x_5, x_7, x_8, x_{10}\right\}, \left\{a_4\right\}\right), \left(\left\{x_2, x_3, x_4, x_5, x_6, x_8, x_9\right\}, \right. \\ \left.\left\{a_3\right\}\right), \left(\left\{x_1, x_3, x_5, x_{10}\right\}, \left\{a_7\right\}\right) \end{array} \right\}.$$

For convenience, the positive feature concept space and the negative feature concept space are represented by $\mathcal{G}_{\mathcal{L}^c\mathcal{H}^c} = \left\{c_1, c_2, \ldots, c_9\right\}$ and $\mathcal{G}_{\bar{\mathcal{L}}^c\bar{\mathcal{H}}^c} = \left\{\bar{c}_1, \bar{c}_2, \ldots, \bar{c}_9\right\}$, respectively.

2.2. Multi-label learning

In MLL, let $\mathcal{X} = \{(x, A(x)) \mid x \in U\}$ be the domain of discourse composed of m-dimensional input examples, where $A_i = \left\{a_1\left(x_i\right), a_2\left(x_i\right), \ldots, a_m\left(x_i\right)\right\}$ is the set of features. The non-empty finite label set $L = \left\{l_1, l_2, \ldots, l_q\right\}$ contains q possible label variables. The possibility label subset of any object $x_i \in U$ in the universe is represented by a vector $L_i = \left\{l_1\left(x_i\right), l_2\left(x_i\right), \ldots, l_q\left(x_i\right)\right\}$ with q-dimensional label values. If the label l_j is related to the object $x_i \in U$, then the label value $l_j\left(x_i\right) = 1$, otherwise the label value $l_j\left(x_i\right) = 0$.

3. Correlation concept-cognitive multi-label learning (3CMLL)

The purpose of MLL is to establish the connection between features and labels. In practice, there will be a certain correlation between multiple features. Similarly, there will be a certain correlation between multiple labels. In addition, there is also a certain correlation between objects and features or labels. Therefore, it is of great significance to explore the correlation between objects, features and labels. One of the important backgrounds for the proposal of concept-cognitive computing is data science, which focuses on acquiring valuable knowledge from research data. Based on the concept-cognitive learning theory, this section conducts cognitive learning on the feature space and the label space to mine the correlation between objects, features and labels. Furthermore, the extent information in the concept structure is used as a bridge to build the feature concepts and the multi-label concepts to obtain the connection between the features and the labels.

3.1. Multi-label feature concept spaces

The cognitive computing of the label space from the perspective of concept cognition is conducive to depicting the knowledge structure between objects and labels. Further referring to the space formed by multi-label concepts as multi-label concept space, the structured hierarchical information of multi-label concept space helps to construct and apply knowledge structures of labels.

Let $\left(U, L, R_L\right)$ be a regular formal context, U and L represent an object set and a label set, respectively. The binary relationship $R_L : U \times L \to \{0, 1\}$ between U and L, that is, $xR_L l = 1$ represents object x with label l. The power sets of U and L are represented by 2^U and 2^L, respectively, then $\mathcal{L}^l : 2^U \to 2^L$ and $\mathcal{H}^l : 2^L \to 2^U$ are two set-valued mappings. In addition, for $\widetilde{X} \subseteq U, \widetilde{L} \subseteq L$, the operators $\mathcal{L}^l$ and $\mathcal{H}^l$ are defined as follows:

$$\mathcal{L}^l\left(\widetilde{X}\right) = \left\{l \in L \mid \forall x \in \widetilde{X}, xR_L l = 1\right\}, \tag{7}$$

$$\mathcal{H}^l\left(\widetilde{L}\right) = \left\{x \in U \mid \forall l \in \widetilde{L}, xR_L l = 1\right\}. \tag{8}$$

If the binary group $\left(\widetilde{X}, \widetilde{L}\right)$ satisfies $\mathcal{L}^l\left(\widetilde{X}\right) = \widetilde{L}$ and $\mathcal{H}^l\left(\widetilde{L}\right) = \widetilde{X}$, then $\left(\widetilde{X}, \widetilde{L}\right)$ is referred to as a multi-label concept. Where $\widetilde{L}$ is the intent of the concept, that is, the maximal set of the labels that all the objects in $\widetilde{X}$ have in common. $\widetilde{X}$ is the extent of the concept, that is, the maximal set of the objects shared by all the labels in $\widetilde{L}$.

In regular formal context $\left(U, L, R_L\right)$, for any $x \in U$ and $l \in L$, $\left(\mathcal{H}^l\mathcal{L}^l(x), \mathcal{L}^l(x)\right)$ and $\left(\mathcal{H}^l(l), \mathcal{L}^l\mathcal{H}^l(l)\right)$ are called multi-label granular concepts under the operators $\mathcal{L}^l$ and $\mathcal{H}^l$. Moreover, we denote the multi-label concept space by, which are the sets of all multi-label granular concepts, that is:

$$\mathcal{G}_{\mathcal{L}^l\mathcal{H}^l} = \left\{\left(\mathcal{H}^l\mathcal{L}^l(x), \mathcal{L}^l(x)\right) \mid x \in U\right\} \cup \left\{\left(\mathcal{H}^l(l), \mathcal{L}^l\mathcal{H}^l(l)\right) \mid l \in L\right\}. \tag{9}$$

Let (U, A, R) and $\left(U, L, R_L\right)$ be two regular formal context, $R : U \times A \to \{0, 1\}$ and $R_L : U \times L \to \{0, 1\}$. Then $\left(U, A, R, L, R_L\right)$ is called a multi-label regular formal context, where A is the feature set and L is the label set.

Example 2 (*Continued with Example 1*). Labels are added to the regular formal context of Table 1. $L = \left\{l_1, l_2, \ldots, l_6\right\}$ is 7 labels, representing vertebrate, invertebrate, mammal, aves, felidae and arachnid, respectively.

Table 2
A multi-label regular formal context.

U	a_1	a_2	a_3	a_4	a_5	a_6	a_7	l_1	l_2	l_3	l_4	l_5	l_6
x_1	1	1	1	0	1	1	0	1	0	0	1	0	0
x_2	1	1	0	0	0	1	1	1	0	1	0	0	0
x_3	1	1	0	1	0	1	0	1	0	1	0	1	0
x_4	0	0	0	0	1	0	1	0	1	0	0	0	0
x_5	0	0	0	0	1	0	0	0	1	0	0	0	0
x_6	1	1	0	1	0	1	1	1	0	1	0	1	0
x_7	1	1	1	0	1	1	1	1	0	0	1	0	0
x_8	0	0	0	0	1	0	1	0	1	0	0	0	1
x_9	1	1	0	1	0	1	1	1	0	1	0	1	0
x_{10}	1	1	1	0	1	1	0	1	0	0	1	0	0

For this multi-label regular formal context in Table 2, the multi-label concept spaces $\mathcal{G}_{\mathcal{L}^l\mathcal{H}^l}$ is shown as follows:

$$\mathcal{G}_{\mathcal{L}^l\mathcal{H}^l}=\left\{\begin{array}{c}\left(\left\{x_1,x_7,x_{10}\right\},\left\{l_1,l_4\right\}\right),\left(\left\{x_2,x_3,x_6,x_9\right\},\left\{l_1,l_3\right\}\right),\\ \left(\left\{x_3,x_6,x_9\right\},\left\{l_1,l_3,l_5\right\}\right),\\ \left(\left\{x_4,x_5,x_8\right\},\left\{l_2\right\}\right),\left(\left\{x_8\right\},\left\{l_2,l_6\right\}\right),\\ \left(\left\{x_1,x_2,x_3,x_6,x_7,x_9,x_{10}\right\},\left\{l_1\right\}\right)\end{array}\right\}.$$

For convenience, the multi-label concept space is represented by $\mathcal{G}_{\mathcal{L}^l\mathcal{H}^l}=\left\{lc_1,lc_2,\dots,lc_6\right\}$.

Taking the two concepts of $\left(\left\{x_2,x_3,x_6,x_9\right\},\left\{l_1,l_3\right\}\right)$ and $\left(\left\{x_3,x_6,x_9\right\},\left\{l_1,l_3,l_5\right\}\right)$ as an example, whale x_2, cat x_3, tiger x_6, and leopard x_9 all have labels: vertebrate l_1 and mammal l_3, indicating that there is a certain correlation between the two labels of l_1 and l_3. In addition, cat x_3, tiger x_6, and leopard x_9 all have labels: vertebrate l_1, mammal l_3, and felidae l_5, indicating that there is a certain correlation between l_1, l_3 and l_5.

By using concept-cognitive learning to perform cognitive computing on label space, the knowledge structure between objects and labels is characterized, and the potential correlation between labels can be obtained.

Definition 3. Let $\left(U,A,R,L,R_L\right)$ be a multi-label regular formal context. For any $(X,B)\in\mathcal{G}_{\mathcal{L}^c\mathcal{H}^c}$, $\left(\overline{X},\overline{B}\right)\in\mathcal{G}_{\bar{\mathcal{L}}^c\bar{\mathcal{H}}^c}$ and $\left(\widetilde{X},\widetilde{L}\right)\in\mathcal{G}_{\mathcal{L}^l\mathcal{H}^l}$, if $X\subseteq\widetilde{X}$, $\overline{X}\subseteq\widetilde{X}$, the positive multi-label feature concept space $\mathcal{S}_{\left(\widetilde{X},\widetilde{L}\right)}$ and the negative multi-label feature concept space $\overline{\mathcal{S}}_{\left(\widetilde{X},\widetilde{L}\right)}$ based on multi-label concept $\left(\widetilde{X},\widetilde{L}\right)$ are defined as:

$$\mathcal{S}_{\left(\widetilde{X},\widetilde{L}\right)}=\left\{(X,B)\,|X\subseteq\widetilde{X},\forall(X,B)\in\mathcal{G}_{\mathcal{L}^c\mathcal{H}^c}\right\},\tag{10}$$

$$\overline{\mathcal{S}}_{\left(\widetilde{X},\widetilde{L}\right)}=\left\{\left(\overline{X},\overline{B}\right)|\overline{X}\subseteq\widetilde{X},\forall\left(\overline{X},\overline{B}\right)\in\mathcal{G}_{\bar{\mathcal{L}}^c\bar{\mathcal{H}}^c}\right\}.\tag{11}$$

The positive multi-label feature concept space $\mathcal{S}_{\left(\widetilde{X},\widetilde{L}\right)}$ is a collection of all positive feature concepts associated with multi-label concept $\left(\widetilde{X},\widetilde{L}\right)$. Similarly, the negative multi-label feature concept space $\overline{\mathcal{S}}_{\left(\widetilde{X},\widetilde{L}\right)}$ is a set of all negative feature concepts associated with multi-label concept $\left(\widetilde{X},\widetilde{L}\right)$. For convenience, we call the positive multi-label feature concept space and the negative multi-label feature concept space as multi-label feature concept space. On the basis of the explanation above, the procedure of constructing multi-label concept space is proposed in Algorithm 1.

Example 3 (*Continued with Example 2*)**.** For multi-label concepts lc_2 and lc_4, their positive multi-label feature concept space and negative multi-label feature concept space are distributed as follows:

$$\mathcal{S}_{lc_2}=\left\{c_3,c_6\right\},\overline{\mathcal{S}}_{lc_2}=\left\{\bar{c}_2,\bar{c}_3,\bar{c}_6\right\};$$
$$\mathcal{S}_{lc_4}=\emptyset,\overline{\mathcal{S}}_{lc_4}=\left\{\bar{c}_4,\bar{c}_5\right\}.$$

Algorithm 1: Constructing the multi-label feature concept spaces.

Input: A multi-label regular formal context $\left(U,A,R,L,R_L\right)$.
Output: the positive and negative multi-label feature concept space $\mathcal{S}_{\left(\widetilde{X},\widetilde{L}\right)}$ and $\overline{\mathcal{S}}_{\left(\widetilde{X},\widetilde{L}\right)}$.

```
for each x ∈ U and a ∈ A do
    Construct a positive feature concept (X, B) by Definition 1;
    G_{L^c H^c} ← (X, B);
    Construct a negative feature concept (X̄, B̄) by Definition 1;
    G_{L̄^c H̄^c} ← (X̄, B̄);
    Construct a multi-label concept (X̃, L̃) by Formula (7) and (8);
    G_{L^l H^l} ← (X̃, L̃);
end
for each (X̃, L̃) ∈ G_{L^l H^l} do
    for each (X, B) ∈ G_{L^c H^c} do
        if X ⊆ X̃ then
            S_(X̃,L̃) ← (X, B)
        end
    end
    for each (X̄, B̄) ∈ G_{L̄^c H̄^c} do
        if X̄ ⊆ X̃ then
            S̄_(X̃,L̃) ← (X̄, B̄)
        end
    end
end
return S_(X̃,L̃) and S̄_(X̃,L̃).
```

It can be seen that the feature concepts in the multi-label feature concept space are related to their corresponding multi-label concepts. When no feature concept can be related to a multi-label concept, its space will be an empty set, such as $\mathcal{S}_{lc_4}=\emptyset$. The operation of multi-label feature concept space as depicted in Fig. 2.

3.2. Correlation concept and correlation concept spaces

In the previous section, the multi-label feature concept space represents the set of all feature concepts associated with the multi-label concept. By taking into account the relationship between the key feature elements and the concepts in the multi-label feature concept space, this section presents and quantifies the association between key feature elements and labels. Based on the structure of the concept, we will consider both the extent and intent relevancy.

Definition 4. Let $\left(U,A,R,L,R_L\right)$ be a multi-label regular formal context and $\mathcal{S}_{\left(\widetilde{X},\widetilde{L}\right)}$ be a positive multi-label feature concept space. Then the positive key feature set F of multi-label concept $\left(\widetilde{X},\widetilde{L}\right)$ is defined as follows:

$$F=\left\{a\left|\frac{\left|\left\{(X,B)\in\mathcal{S}_{\left(\widetilde{X},\widetilde{L}\right)}|a\in B\right\}\right|}{\left|\mathcal{S}_{\left(\widetilde{X},\widetilde{L}\right)}\right|}\geqslant\beta_1\right.\right\},\tag{12}$$

where, $\left|\left\{(X,B)\in\mathcal{S}_{\left(\widetilde{X},\widetilde{L}\right)}|a\in B\right\}\right|$ denotes the number of concepts with feature a in $\mathcal{S}_{\left(\widetilde{X},\widetilde{L}\right)}$, and $\left|\mathcal{S}_{\left(\widetilde{X},\widetilde{L}\right)}\right|$ denotes the number of all

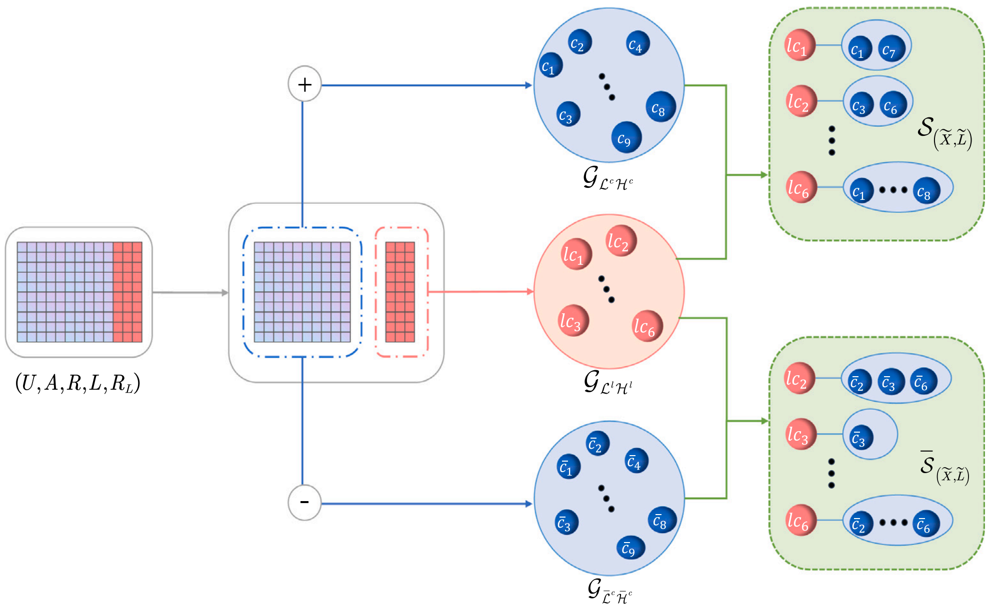


Fig. 2. The construction process of the multi-label feature concept spaces.

concepts in $\mathcal{S}_{(\tilde{X},\tilde{L})}$. The positive key feature set F represents the feature set whose proportion of the number of concepts with feature a in $\mathcal{S}_{(\tilde{X},\tilde{L})}$ to the total number of concepts in $\mathcal{S}_{(\tilde{X},\tilde{L})}$ above or equal β_1. In particular, when $\beta_1 = 1$, it means that all positive feature concepts in $\mathcal{S}_{(\tilde{X},\tilde{L})}$ have feature a.

Similarly, for the multi-label concept $\left(\tilde{X}, \tilde{L}\right)$, the negative key feature set $\overline{F}$ is defined as follows:

$$\overline{F} = \left\{ a \left| \frac{\left|\left\{\left(\overline{X}, \overline{B}\right) \in \overline{\mathcal{S}}_{(\tilde{X},\tilde{L})} | a \in \overline{B}\right\}\right|}{\left|\overline{\mathcal{S}}_{(\tilde{X},\tilde{L})}\right|} \geqslant \beta_2 \right.\right\}. \tag{13}$$

The negative key feature set $\overline{F}$ represents the feature set whose proportion of the number of concepts with feature a in $\overline{\mathcal{S}}_{(\tilde{X},\tilde{L})}$ to the total number of concepts in $\overline{\mathcal{S}}_{(\tilde{X},\tilde{L})}$ above or equal β_2. In particular, when $\beta_2 = 1$, it means that all negative feature concepts in $\overline{\mathcal{S}}_{(\tilde{X},\tilde{L})}$ have feature a.

Definition 5. Let (U, A, R, L, R_L) be a multi-label regular formal context and $\mathcal{S}_{(\tilde{X},\tilde{L})}$ be a positive multi-label feature concept space. For a multi-label concept $\left(\tilde{X}, \tilde{L}\right)$ and its positive key feature set F, the positive intent relevancy is defined as:

$$r_{in} = \frac{1}{s}\sum_{k=1}^{s} \frac{\left|\left\{(X, B) \in \mathcal{S}_{(\tilde{X},\tilde{L})} | a_k \in B \cap F\right\}\right|}{\left|\mathcal{S}_{(\tilde{X},\tilde{L})}\right|}, \tag{14}$$

where s is the number of positive key feature elements in the positive key feature set F, and $\left|\left\{(X, B) \in \mathcal{S}_{(\tilde{X},\tilde{L})} | a_k \in B \cap F\right\}\right|$ is the number of the positive feature concepts with positive key feature a_k in $\mathcal{S}_{(\tilde{X},\tilde{L})}$.

Positive intent relevancy r_{in} reflects the correlation measure between the feature concept intent B and the positive key feature set F in the positive multi-label feature concept space $\mathcal{S}_{(\tilde{X},\tilde{L})}$. When r_{in} is larger, it means that the correlation between the concept intent B and the positive key feature set F is greater. In particular, when all concepts in $\mathcal{S}_{(\tilde{X},\tilde{L})}$ have the positive key features a_k in F, then $r_{in} = 1$.

Similarly, for the multi-label concept $\left(\tilde{X}, \tilde{L}\right)$ and its negative key feature set $\overline{F}$, the negative intent relevancy is defined as:

$$\bar{r}_{in} = \frac{1}{s}\sum_{k=1}^{s} \frac{\left|\left\{\left(\overline{X}, \overline{B}\right) \in \overline{\mathcal{S}}_{(\tilde{X},\tilde{L})} | a_k \in \overline{B} \cap \overline{F}\right\}\right|}{\left|\overline{\mathcal{S}}_{(\tilde{X},\tilde{L})}\right|}, \tag{15}$$

where s is the number of negative key feature elements in the negative key feature set $\overline{F}$, and $\left\{\left(\overline{X}, \overline{B}\right) \in \overline{\mathcal{S}}_{(\tilde{X},\tilde{L})} | a_k \in \overline{B} \cap \overline{F}\right\}$ is the number of negative feature concepts with negative key feature a_k in $\overline{\mathcal{S}}_{(\tilde{X},\tilde{L})}$.

Negative intent relevancy $\bar{r}_{in}$ reflects the correlation measure between the feature concept intent $\overline{B}$ and the negative key feature set $\overline{F}$ in the negative multi-label feature concept space $\overline{\mathcal{S}}_{(\tilde{X},\tilde{L})}$. When $\bar{r}_{in}$ is larger, it means that the correlation between the concept intent $\overline{B}$ and the negative key feature set $\overline{F}$ is greater. In particular, when all concepts in $\overline{\mathcal{S}}_{(\tilde{X},\tilde{L})}$ have the negative key features a_k in $\overline{F}$, then $\bar{r}_{in} = 1$.

Definition 6. Let (U, A, R, L, R_L) be a multi-label regular formal context and $\mathcal{S}_{(\tilde{X},\tilde{L})}$ be a positive multi-label-feature concept space. For the multi-label concept $\left(\tilde{X}, \tilde{L}\right)$ and its positive key feature set F, any $(X, B) \in \mathcal{S}_{(\tilde{X},\tilde{L})}$, if $B \cap F \neq \emptyset$, then the positive extent relevancy is defined as:

$$r_{ex} = \frac{|\cup X|}{\left|\tilde{X}\right|}, \tag{16}$$

where, $|\cup X|$ is the number of objects in all positive feature concepts extent with positive key feature elements in $\mathcal{S}_{(\tilde{X},\tilde{L})}$, and $\left|\tilde{X}\right|$ is the number of objects in the multi-label concept extent.

The positive extent relevancy r_{ex} reflects the correlation measure between the positive feature concepts with the positive key feature elements and the multi-label concept. When r_{ex} is larger, it means that the correlation between the multi-label concept $\left(\tilde{X}, \tilde{L}\right)$ and the positive feature concepts (X, B) is greater.

Similarly, for the multi-label concept $\left(\widetilde{X},\widetilde{L}\right)$ and its negative key feature set $\overline{F}$, any $\left(\overline{X},\overline{B}\right)\in\overline{S}_{\left(\widetilde{X},\widetilde{L}\right)}$, if $\overline{B}\cap\overline{F}\neq\emptyset$, the negative extent relevancy is defined as:

$$\bar{r}_{ex}=\frac{\left|\cup\overline{X}\right|}{\left|\widetilde{X}\right|},\tag{17}$$

where, $\left|\cup\overline{X}\right|$ is the number of objects in all negative feature concepts extent with negative key feature elements in $\overline{S}_{\left(\widetilde{X},\widetilde{L}\right)}$, and $\left|\widetilde{X}\right|$ is the number of objects in the multi-label concept extent.

The negative extent relevancy $\bar{r}_{ex}$ reflects the correlation measure between the negative feature concepts with the negative key feature elements and the label concept. When $\bar{r}_{ex}$ is larger, it means that the correlation between the multi-label concept $\left(\widetilde{X},\widetilde{L}\right)$ and the negative feature concepts $\left(\overline{X},\overline{B}\right)$ is greater.

Therefore, the positive relevancy r and the negative relevancy $\bar{r}$ can be calculated as follows:

$$r=\frac{r_{in}+r_{ex}}{2},\tag{18}$$

$$\bar{r}=\frac{\bar{r}_{in}+\bar{r}_{ex}}{2}.\tag{19}$$

The positive relevancy r represents the positive correlation measure between the object set $\widetilde{X}$, the label set $\widetilde{L}$ and the positive key feature set F. When the positive relevancy r is larger, it means that the correlation measure between the object set $\widetilde{X}$, the label set $\widetilde{L}$ and the positive key feature set F is larger. Similarly, The negative relevancy $\bar{r}$ represents the negative correlation measure between the object set $\widetilde{X}$, the label set $\widetilde{L}$ and the negative key feature set $\overline{F}$. When the negative relevancy $\bar{r}$ is larger, it means that the negative correlation measure between the object set $\widetilde{X}$, the label set $\widetilde{L}$ and the negative key feature set $\overline{F}$ is larger.

Definition 7. Let (U,A,R,L,R_L) be a multi-label regular formal context. For multi-label concept $\left(\widetilde{X},\widetilde{L}\right)$, its positive key feature set F and negative key feature set $\overline{F}$, positive relevancy r and negative relevancy $\bar{r}$, then $\left(\widetilde{X},\widetilde{L},F,r\right)$ and $\left(\widetilde{X},\widetilde{L},\overline{F},\bar{r}\right)$ are called positive and negative correlation concept, respectively.

Definition 8. Let (U,A,R,L,R_L) be a multi-label regular formal context, $\left(\widetilde{X},\widetilde{L},F,r\right)$ be a positive correlation concept, and $\left(\widetilde{X},\widetilde{L},\overline{F},\bar{r}\right)$ be a negative correlation concept. When the correlation degree more than threshold α, then the positive and negative correlation concept space are defined as follows, respectively:

$$CCS^{\alpha}=\left\{\left(\widetilde{X},\widetilde{L},F,r\right)\mid\left(\widetilde{X},\widetilde{L}\right)\in\mathcal{G}_{\mathcal{L}^l\mathcal{H}^l},r\geqslant\alpha\right\},\tag{20}$$

$$\overline{CCS}^{\alpha}=\left\{\left(\widetilde{X},\widetilde{L},\overline{F},\bar{r}\right)\mid\left(\widetilde{X},\widetilde{L}\right)\in\mathcal{G}_{\mathcal{L}^l\mathcal{H}^l},\bar{r}\geqslant\alpha\right\}.\tag{21}$$

The positive correlation concept space CCS^{α} and the negative correlation concept space $\overline{CCS}^{\alpha}$ represent the set of all positive and negative correlation concepts which correlation degree reaches α. The specific process of constructing the positive and negative correlation concept spaces is shown in algorithm 2.

Example 4 (*Continued with Example 3*)**.** If the threshold $\alpha=0.6$ and $\beta_1=\beta_2=0.5$. The positive correlation concept space CCS^{α} and the negative correlation concept space $\overline{CCS}^{\alpha}$ are calculated as follows:

$$CCS^{0.6}=\left\{\begin{array}{c}\left(\{x_1,x_7,x_{10}\},\{l_1,l_4\},\{a_1,a_2,a_3,a_5,a_6,a_7\},0.9583\right),\\\left(\{x_2,x_3,x_6,x_9\},\{l_1,l_3\},\{a_1,a_2,a_4,a_6,a_7\},0.8250\right),\\\left(\{x_3,x_6,x_9\},\{l_1,l_3,l_5\},\{a_1,a_2,a_4,a_6,a_7\},0.95\right),\\\left(\{x_1,x_2,x_3,x_6,x_7,x_9,x_{10}\},\{l_1\},\{a_1,a_2,a_6,a_7\},0.9375\right)\end{array}\right\},$$

$$\overline{CCS}^{0.6}=\left\{\begin{array}{c}\left(\{x_2,x_3,x_6,x_9\},\{l_1,l_3\},\{a_3,a_5\},1\right),\\\left(\{x_3,x_6,x_9\},\{l_1,l_3,l_5\},\{a_3,a_5,a_7\},0.6667\right),\\\left(\{x_4,x_5,x_8\},\{l_2\},\{a_1,a_2,a_3,a_4,a_6,a_7\},0.9583\right),\\\left(\{x_1,x_2,x_3,x_6,x_7,x_9,x_{10}\},\{l_1\},\{a_3,a_5\},0.7857\right)\end{array}\right\}.$$

For convenience, the positive correlation concept space and the negative correlation concept space are represented by $CCS^{0.6}=\{rc_1,rc_2,\dots,rc_4\}$ and $\overline{CCS}^{0.6}=\{\overline{rc}_1,\overline{rc}_2,\dots,\overline{rc}_4\}$, respectively.

Taking $\left(\{x_1,x_7,x_{10}\},\{l_1,l_4\}\right)$ as an example, the operation of positive correlation concept and the positive correlation concept space as depicted in Fig. 3.

Fig. 3 serves as an example to illustrate the process of constructing the positive correlation concept and correlation concept space. By utilizing Definition 5, the extent relevancy (r_{ex}) measures the correlation between feature concepts and multi-label concepts from an extent perspective. Simultaneously, Definition 4 quantifies the intent relevancy (r_{in}) by measuring the correlation between feature concepts and key features from an intent perspective. These two measures are combined in a comprehensive relevancy (r), thereby defining the correlation concept $\left(\widetilde{X},\widetilde{L},F,r\right)$. To enable effective concept prediction, concepts with relevancy (r) greater than or equal to α are selected to the correlation concept space (CCS^{α}). Subsequently, CCS^{α} will be utilized for label prediction.

Algorithm 2: The construction of positive correlation concept spaces.

Input: The positive multi-label feature concept space $S_{\left(\widetilde{X},\widetilde{L}\right)}$, threshold β_1 and α.
Output: The positive correlation concept space CCS^{α}.

```
for each a ∈ B and (X, B) ∈ S_(X̃,L̃) do
    if |{(X,B)∈S_(X̃,L̃) | a∈B}| / |S_(X̃,L̃)| ⩾ β1 then
        F ← a;
        Compute the positive intent relevancy r_in by Formula (14);
    end
    if B ∩ F ≠ ∅ then
        Compute the positive extent relevancy r_ex by Formula (16);
        Compute the positive relevancy r by Formula (18);
        Get the positive correlation concept (X̃, L̃, F, r);
        if r ⩾ α then
            CCS^α ← (X̃, L̃, F, r);
        end
    end
end
return CCS^α.
```

3.3. Labels prediction

Definition 9. Let (U,A,R,L,R_L) be a multi-label regular formal context, where $A=\{a_1,a_2,\dots,a_m\}$ and $L=\{l_1,l_2,\dots,l_q\}$, $\left(X_j,L_j,F_j,r\right)\in CCS^{\alpha}$ be the positive correlation concept. If $a_i\in A$, $l_j\in L$, the total positive correlation value P_{ij} between a_i and l_j is defined as follows:

$$P_{ij}=\sum_{\left(\widetilde{X},\widetilde{L},F,r\right)\in CCS^{\alpha}}p_{ij},$$

where $p_{ij}=\begin{cases}r & ,if\ a_i\in F\ and\ l_j\in\widetilde{L}\\0 & ,otherwise\end{cases}$, P_{ij} denotes the sum of the correlation degree of all positive correlation concepts in the positive correlation concept space with respect to feature a_i and label l_j.

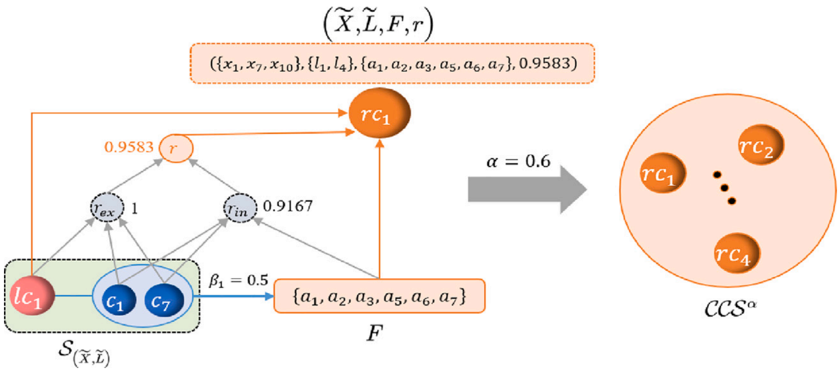


Fig. 3. The operation of the positive correlation concept and the positive correlation concept space.

After normalization, we can obtain:

$$P'_{ij} = \frac{P_{ij}}{\sum_{i=1}^{m}\sum_{j=1}^{q} P_{ij}}.$$

Therefore, the positive related feature label matrix can be expressed as:

$$FL^{+} = \left(P'_{ij}\right)_{m\times q} = \begin{array}{c} \\ a_1 \\ a_2 \\ \vdots \\ a_m \end{array}\begin{array}{c} \begin{array}{cccc} l_1 & l_2 & \cdots & l_q \end{array} \\ \begin{pmatrix} P'_{11} & P'_{12} & \cdots & P'_{1q} \\ P'_{21} & P'_{22} & \cdots & P'_{2q} \\ \vdots & \vdots & \ddots & \vdots \\ P'_{m1} & P'_{m2} & \cdots & P'_{mq} \end{pmatrix} \end{array}.$$

In the positive related feature label matrix, the value of P'_{ij} reflects the total positive correlation between feature a_i and label l_j. If $P'_{ij} = 0$, it indicates that the object with feature a_i will not have a label l_j. When P'_{ij} is larger, it means that if the object has feature a_i, the possibility of having label l_j is greater. In addition, we use $B_j^{+} = \left\{a_i | P'_{ij} > 0\right\}$ to represent the feature set that is positive correlated with the label l_j.

Similarly, Let $\left(\widetilde{X},\widetilde{L},\overline{F},\bar{r}\right) \in \overline{CCS}^{\alpha}$ be the negative correlation concept. If $a_i \in A$, $l_j \in L$, the total negative correlation value $\overline{P}_{ij}$ between a_i and l_j is defined as follows:

$$\overline{P}_{ij} = \sum_{\left(\widetilde{X},\widetilde{L},\overline{F},\bar{r}\right)\in\overline{CCS}^{\alpha}} \bar{p}_{ij},$$

where $\bar{p}_{ij} = \begin{cases} \bar{r} & ,if\ a_i \in \overline{F}\ and\ l_j \in \widetilde{L} \\ 0 & ,otherwise \end{cases}$, $\overline{P}_{ij}$ denotes the sum of the correlations of all negative correlation concepts in the negative correlation concept space with respect to feature a_i and label l_j.

After normalization, we can obtain:

$$\overline{P}'_{ij} = \frac{\overline{P}_{ij}}{\sum_{i=1}^{m}\sum_{j=1}^{q} \overline{P}_{ij}}.$$

Therefore, the negative related feature label matrix can be expressed as:

$$FL^{-} = \left(\overline{P}'_{ij}\right)_{m\times q} = \begin{array}{c} \\ a_1 \\ a_2 \\ \vdots \\ a_m \end{array}\begin{array}{c} \begin{array}{cccc} l_1 & l_2 & \cdots & l_q \end{array} \\ \begin{pmatrix} \overline{P}'_{11} & \overline{P}'_{12} & \cdots & \overline{P}'_{1q} \\ \overline{P}'_{21} & \overline{P}'_{22} & \cdots & \overline{P}'_{2q} \\ \vdots & \vdots & \ddots & \vdots \\ \overline{P}'_{m1} & \overline{P}'_{m2} & \cdots & \overline{P}'_{mq} \end{pmatrix} \end{array}.$$

In the negative related feature label matrix, the value of $\overline{P}'_{ij}$ reflects the total negative correlation between feature a_i and label l_j. If $\overline{P}'_{ij} = 0$, it indicates that the object does not have feature a_i and will not have a label l_j. When $\overline{P}'_{ij}$ is larger, it means that if the object do not have feature a_i, the possibility of having label l_j is greater. In addition, we use $B_j^{-} = \left\{a_i | \overline{P}'_{ij} > 0\right\}$ to represent the feature set that is negative correlated with the label l_j.

Definition 10. Let $\left(U, A, R, L, R_L\right)$ be a multi-label regular formal context, where $A = \left\{a_1, a_2, \dots, a_m\right\}$ and $L = \left\{l_1, l_2, \dots, l_q\right\}$. FL^{+} and FL^{-} are positive and negative related feature label matrices, B_j^{+} and B_j^{-} are feature sets that have positive and negative correlations with the label l_k, respectively. x is the prediction object with feature set B. Then the predicted label value $pl_j(x)$ of x about label l_j is defined as follows:

$$pl_j(x) = \sum_{i=1}^{m} b_{ij}, \tag{22}$$

where $b_{ij} = \begin{cases} \frac{P'_{ij}}{\left|B_j^{+}\cup B\right|}, & a_i \in B \\ \frac{\overline{P}'_{ij}}{\left|B_j^{-}\cup B\right|}, & a_i \notin B \end{cases}$, $|\cdot|$ represents the element count in the set.

Furthermore, the predicted label vector of x for all labels is:

$$PL(x) = \left(pl_1(x) \quad pl_2(x) \quad \cdots \quad pl_q(x)\right).$$

$PL(x)$ reflects the predicted values for all the labels of x, the predicted label value $pl_j(x)$ indicates the possibility that x has label l_j. The larger the value of $pl_j(x)$, the greater the possibility that the object x has the label l_j. Based on the above theory, the algorithm 3 is label prediction when new objects are randomly added.

Algorithm 3: Labels prediction based on the positive and negative correlation concepts.

Input: The positive correlation concept space CCS^{α} and the negative correlation concept space $\overline{CCS}^{\alpha}$, the newly added object x.
Output: The predicted label vector of x.

for *each* $a_i \in A$ *and* $l_j \in L$ **do**
  Compute the positive related feature label matrix $FL^{+} = \left(P'_{ij}\right)_{m\times q}$;
  Compute the negative related feature label matrix $FL^{-} = \left(\overline{P}'_{ij}\right)_{m\times q}$;
  Compute the predicted label value $pl_j(x)$;
  $PL(x) \leftarrow pl_j(x)$;
end
return $PL(x) = \left(pl_1(x) \quad pl_2(x) \quad \cdots \quad pl_q(x)\right)$.

As shown in Fig. 4, the process of 3CMLL consists of three main parts: (1) Constructing multi-label feature concept spaces; (2) Constructing the correlation concept spaces; (3) Form related feature label matrix. For ease of understanding, our framework is represented on

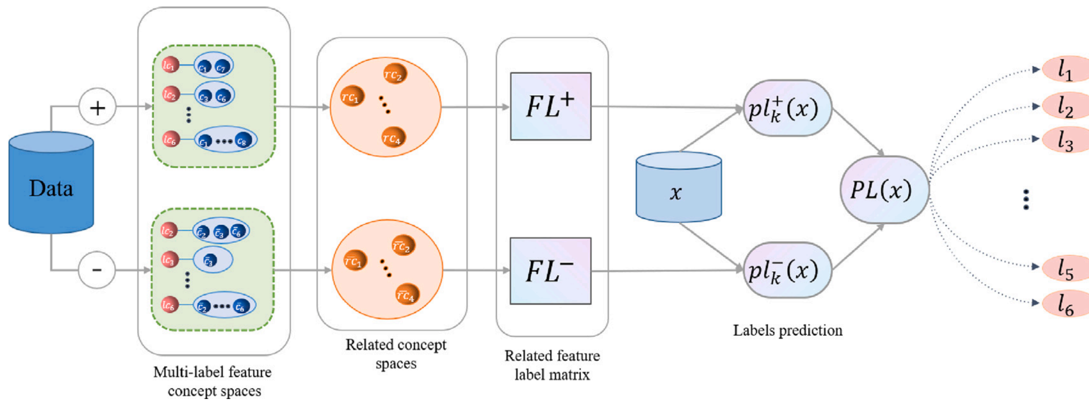


Fig. 4. Overview of our proposed 3CMLL process.

the basis of Example 4. In the first part, the multi-label feature space is constructed by associating positive and negative feature concepts with multi-label concepts. The second part is to generate correlation concepts and construct correlation concept space from two perspectives of extent relevancy and intent relevancy. In the third part, the related feature label matrix is constructed, and the predicted values of each label of the new instance x are calculated. Finally, the possible values of each label are output by predicting the label vector $PL(x)$.

The time complexity of 3CMLL has been analyzed in this article. As described from Algorithm 1, it primarily consists of two parts: the process of acquiring concepts by cognitive computation and the construction of a multi-label feature concept space. Typically, when $m, q \ll n$ in the multi-label data, the time complexity for acquiring concepts is $O(n)$. Assuming $\left|\mathcal{G}_{\mathcal{L}^c\mathcal{H}^c}\right| \geqslant \left|\mathcal{G}_{\tilde{\mathcal{L}}^c\tilde{\mathcal{H}}^c}\right|$, the time complexity for constructing the multi-label feature concept space is $O\left(\left|\mathcal{G}_{\mathcal{L}^c\mathcal{H}^c}\right|\left|\mathcal{G}_{\mathcal{L}^l\mathcal{H}^l}\right|\right)$. Owing to the similarity of constructing positive and negative correlation concept spaces, the time complexity of Algorithm 2 is $O\left(m\left|S_{(\tilde{X},\tilde{L})}\right|\right)$. Algorithm 3 is the process of label prediction, and its time complexity is $O(mq)$. Hence, the overall time complexity of 3CMLL is $O\left(n+\left|\mathcal{G}_{\mathcal{L}^c\mathcal{H}^c}\right|\left|\mathcal{G}_{\mathcal{L}^l\mathcal{H}^l}\right|+m\left|S_{(\tilde{X},\tilde{L})}\right|+mq\right)$.

4. Experiments

In order to assess 3CMLL, we run some experiments in this part. Table 3 shows the detailed information of 10 multi-label datasets, including sample quantity, feature quantity, label quantity, and domain. These datasets cover many types, such as text, music, image, medicine and biology, which can be downloaded for free from Multi-Label Classification Dataset Repository and Yahoo Web Pages. The datasets in Table 3 are not regular formal contexts, so the datasets need to be preprocessed before the experiments. Through the commonly used binarization method, the numerical features are converted to 0 or 1, which satisfies the definition of regular formal context.

4.1. Comparison methods and experimental evaluation metrics

We contrast classic multi-label algorithm CC [40] and the following state-of-the-art multi-label algorithms with 3CMLL. LLSF-DL [46]: Parameters α, β, γ are searched in $\{4^{-5}, 4^{-4}, \dots, 4^{5}\}$, and ρ is searched in $\{0.1, 1, 10\}$. MSWL [47]: The parameters are set to $\alpha = 0.1, \beta = 0.1, \gamma = 1$. GLMAM [48]: The neighbor number is set to 10 and the parameters $\lambda_1 = 10^2, \lambda_2 = 10^{-2}, \lambda_3 = 10^{-2}$ and $g = 9$. GLFS [49]: The parameters are $\alpha = 0.2, \beta = 10^3, \gamma = 10$ and $\lambda = 0.4$. 2SML [50]: The parameters are set to $\lambda_1 = 10^{-3}, \lambda_2 = 10^{-3}, \lambda_3 = 10^{-4}$ and $\alpha = 0.6$.

We will assess the algorithm's performance using the four multi-label classification evaluation metrics that are frequently used, namely Average Precision, Coverage, One Error and Ranking Loss. Suppose we are given with a test set $\mathcal{T} = \{(x_i, L_i) \mid 1 \leqslant i \leqslant s\}$ and all labels are arranged in descending order $\{f_1(x_i), f_2(x_i), \dots, f_q(x_i)\}$. The specific calculation method of each evaluation metrics is as follows:

Average Precision (AP): Let $R_i = \{l_j \mid rank(x_i, l_j) \leqslant rank(x_i, l_k)\}$, for any object $x_i \in U$, this index is used to measure the average probability of other related labels ranking before the selected related label $l_j \in L_i$. The multi-label classification algorithm performs better the larger the value of the evaluation metric.

$$Average\ Precision = \frac{1}{s}\sum_{i=1}^{s}\frac{1}{|L_i|}\sum_{l_j, l_k \in L_i}\frac{|R_i|}{rank(x_i, l_k)}.$$

Coverage(CV): In the label ranking list, it takes an average of how many steps to move down to cover all the real tag values contained in an object. The classification performance improves as the metric's value decreases.

$$Coverage = \frac{1}{q}\left(\frac{1}{s}\sum_{i=1}^{s}\max_{l_k \in L_i} rank(x_i, l_k) - 1\right).$$

One-error(OE): Evaluate the proportion of objects whose top-ranked labels are not in the relevant label set. The performance of multi-label classification improves with decreasing metric value.

$$One-error = \frac{1}{s}\sum_{i=1}^{s}\mathcal{I}\left(\arg\max_{l \in L} f(x_i) \notin L_i\right).$$

Ranking Loss (RL): Calculate the proportion of all sample reverse sorting label pairs, that is, the ranking of irrelevant labels is higher than that of relevant labels. The performance of the multi-label classification method improves with decreasing metric values.

$$Ranking\ Loss = \frac{1}{s}\sum_{i=1}^{s}\frac{\left|\left\{(l_k, l_j) \mid f_j(x_i) \geqslant f_k(x_i), (l_k, l_j) \in L_i \times \overline{L_i}\right\}\right|}{|L_i|\left|\overline{L_i}\right|},$$

where $\overline{L_i}$ is the complement of set L_i with respect to the label set L.

Table 3
Description of the multi-label datasets.

Data set	Instance	Feature	Label	Cardinality	Domain
CHD49	555	49	6	2.580	Medicine
Genbase	662	1186	27	1.252	Biology
Emotions	593	72	6	1.869	Music
Business	5000	438	30	1.588	Text
Flags	194	19	7	3.392	Image
Image	2000	294	5	1.236	Image
CAL500	502	68	174	26.044	Music
Enron	1702	1001	53	3.378	Text
Scene	2407	294	6	1.074	Image
Computers	5000	681	33	1.508	Text

Table 4
Comparison results of 3CMLL with different correlations.

Method	Average precision (↑)									
	CHD49	Genbase	Emotions	Business	Flags	Image	CAL500	Enron	Scene	Computers
3CMLL-ex	0.7858	0.9637	0.5448	0.8456	0.7725	0.5157	0.4930	0.5359	0.5085	0.5756
3CMLL-in	0.7726	0.9633	0.6107	0.7641	0.7745	0.5546	0.4950	0.5219	0.5003	**0.5777**
3CMLL	**0.7884**	**0.9654**	**0.7509**	**0.8516**	**0.7847**	**0.8163**	**0.4953**	**0.5391**	**0.6159**	0.5758
Method	Coverage (↓)									
	CHD49	Genbase	Emotions	Business	Flags	Image	CAL500	Enron	Scene	Computers
3CMLL-ex	0.4964	0.0150	0.5203	0.1073	0.5820	0.4431	0.7550	**0.2825**	0.3393	**0.1754**
3CMLL-in	0.5075	0.0154	0.4863	0.1115	0.5789	0.3995	**0.7515**	0.2869	0.3560	0.1828
3CMLL	**0.4946**	**0.0147**	**0.3525**	**0.0972**	**0.5598**	**0.1702**	0.7539	0.2833	**0.2961**	0.1788
Method	One-error (↓)									
	CHD49	Genbase	Emotions	Business	Flags	Image	CAL500	Enron	Scene	Computers
3CMLL-ex	**0.2137**	0.0649	0.6068	**0.1350**	0.2137	0.6970	0.1156	0.3503	0.7166	0.4790
3CMLL-in	0.2354	0.0649	0.5377	0.2964	0.2130	0.6475	0.1157	0.3949	0.6980	0.4798
3CMLL	0.2248	**0.0619**	**0.3339**	0.1352	**0.2077**	**0.2945**	**0.1155**	**0.3291**	**0.5335**	**0.4782**
Method	Ranking Loss (↓)									
	CHD49	Genbase	Emotions	Business	Flags	Image	CAL500	Enron	Scene	Computers
3CMLL-ex	0.2332	0.0039	0.2221	0.0555	0.2502	0.4422	0.1830	0.1122	0.3899	**0.1137**
3CMLL-in	0.2392	0.0042	0.3774	0.0632	0.2488	0.4366	**0.1820**	0.1137	0.4102	0.1182
3CMLL	**0.2260**	**0.0037**	**0.2190**	**0.0509**	**0.2306**	**0.1456**	0.1827	**0.1112**	**0.3380**	0.1154

Table 5
Comparison results of 3CMLL with other algorithms for the AP (↑) metric.

Data set	LLSF-DL	MSWL	GLMAM	GLFS	2SML	CC	3CMLL
CHD49	0.7806	0.7811	0.6727	0.7776	0.7815	0.7639	**0.7884**
Genbase	0.9895	0.9615	0.7088	0.9571	0.6231	**0.9962**	0.9654
Emotions	0.7224	0.7465	0.6053	0.6739	0.7483	0.6862	**0.7509**
Business	0.8267	0.8392	0.5451	0.8506	**0.8884**	0.8086	0.8516
Flags	0.5066	0.7782	0.6688	0.7611	0.7818	0.7762	**0.7847**
Image	0.4857	0.6783	0.5406	0.6193	0.6948	0.5838	**0.8163**
CAL500	0.4950	0.4604	0.2727	0.4909	0.4630	0.3893	**0.4953**
Enron	0.5389	0.4934	0.1782	0.5267	0.5337	0.4812	**0.5391**
Scene	0.4292	0.7085	0.4901	0.5841	**0.7373**	0.6107	0.6159
Computers	0.4422	**0.6338**	0.4342	0.6158	0.5529	0.5589	0.5758
Average	0.6217	0.7081	0.5117	0.6857	0.6805	0.6655	**0.7184**
Ave. Rank.	4.6	3.3	6.6	4.2	2.9	4.7	**1.7**

4.2. Experimental results and analysis

In this part, we used five-fold cross-validation for experiments. Specifically, we randomly divide each dataset into five parts, each of which is retained in turn for testing, and the remaining data is merged as a training set. Finally, five results are obtained to calculate the average value. Tables 4–8 show the experimental results. The optimal results for each dataset are displayed in bold. The average and average ranking (Ave. Rank.) results on all datasets are shown in the last two rows.

In Section 3.2, it becomes apparent that the proposed correlation is investigated from both extent and intent aspects, resulting in a more comprehensive understanding of the inherent correlation information within each concept. To verify the importance of studying correlation from both extent and intent perspectives, we will compare the performance of the original model (3CMLL) with the performance of using only extent (3CMLL-ex) or intent (3CMLL-in) relevancy on ten datasets. As Table 4 demonstrates, in most cases, the performance of 3CMLL is better than 3CMLL-ex and 3CMLL-in.

According to the experimental results on 10 multi-label datasets (Tables 5–8), we can draw the following conclusions: 3CMLL is dominant on AP, CV and RL metrics on most datasets, and it performs best both in terms of average and average ranking. Compared with other metrics, 3CMLL performs slightly lower on the OE metric. Although 3CMLL does not perform best on the OE metric, there is no significant difference between 3CMLL and the best performing 2SML in terms of average and average ranking. Indeed, 3CMLL outperforms LLSF-DL, MSWL, GLMAM, GLFS, 2SML and CC on AP, CV and RL metrics. 3CMLL is slightly worse than 2SML on OE metric, but better than LLSF-DL, MSWL, GLMAM, GLFS and CC. Overall, it shows that 3CMLL has certain superiorities over other popular algorithms.

In addition, the Friedman test [51] was used to analyze the statistical significance of the comparison algorithm. Table 9 displays the Friedman statistics F_F and the accompanying critical value for each assessment metric. The initial hypothesis that all comparison approaches perform equivalently is categorically rejected in terms of each evaluation metric, as shown in Table 9, when the significance level $\alpha = 0.05$ is used.

Therefore, we compare and examine the relative performance of the various algorithms using post-hoc test [51]. Nemenyi test [38,51] is used to test whether 3CMLL is significantly competitive with other algorithms. If the corresponding average ranks of two algorithms differ by at least the critical value $CD = q_\alpha \sqrt{\frac{k(k+1)}{6N}}$, the performance of the algorithms will be noticeably different. For Nemenyi test, $q_\alpha = 2.949$ at level $\alpha = 0.05$, and thus $CD = 2.8490\,(k = 7, N = 10)$. The CD diagram of each evaluation metric is shown in Fig. 5, any algorithms with no connections between them are thought to have significantly different performance from one another. It can be found from Fig. 5 that the 3CMLL method is significantly better than GLMAM, LLSF-DL and CC in the evaluation metric AP and RL, and has no significant difference with GLFS, MSWL and 2SML. On the evaluation metric CV, 3CMLL is significantly better than LLSF-DL and GLMAM. In addition, 3CMLL is significantly better than GLMAM and CC for the evaluation metric OE.

In order to visually highlight the differences between other algorithms and 3CMLL, we visualized the ranking of all algorithms on each

Table 6
Comparison results of 3CMLL with other algorithms for the CV (↓) metric.

Data set	LLSF-DL	MSWL	GLMAM	GLFS	2SML	CC	3CMLL
CHD49	0.4493	0.4967	0.6136	0.4484	**0.4472**	0.4879	0.4946
Genbase	0.0148	0.0240	0.1910	0.0219	0.1014	**0.0125**	0.0147
Emotions	0.3574	0.3539	0.4833	0.3873	0.3548	0.3780	**0.3525**
Business	0.1848	0.1136	0.3697	0.0975	**0.0771**	0.1473	0.0972
Flags	0.6172	0.5631	0.7005	**0.5180**	0.5600	0.5632	0.5598
Image	0.4521	0.2718	0.3991	0.2994	0.2560	0.3580	**0.1702**
CAL500	0.7562	0.8427	0.9707	0.7549	0.8700	0.9031	**0.7539**
Enron	0.4739	0.4733	0.7183	0.2941	0.3706	0.5212	**0.2833**
Scene	0.4476	0.1782	0.3656	0.2640	**0.1573**	0.2385	0.2961
Computers	0.4422	0.1872	0.3901	**0.1378**	0.1793	0.2546	0.1788
Average	0.4195	0.3504	0.5202	0.3223	0.3374	0.3201	**0.3201**
Ave. Rank.	5.1	3.8	6.7	2.9	2.8	4.5	**2.2**

Table 7
Comparison results of 3CMLL with other algorithms for the OE (↓) metric.

Data set	LLSF-DL	MSWL	GLMAM	GLFS	2SML	CC	3CMLL
CHD49	0.2920	0.2683	0.3754	0.2636	0.2763	0.3164	**0.2248**
Genbase	0.0030	0.0482	0.3050	0.0534	0.0016	**0.0015**	0.0619
Emotions	0.4216	0.3645	0.5305	0.4712	0.3478	0.4793	**0.3339**
Business	0.1376	0.1631	0.4626	0.1446	**0.1149**	0.1784	0.1352
Flags	**0.0667**	0.2545	0.3640	0.3211	0.2743	0.2757	0.2077
Image	0.6190	0.4969	0.6790	0.6025	0.4767	0.6200	**0.2945**
CAL500	0.2310	0.2584	0.5054	**0.0640**	0.2099	0.3520	0.1155
Enron	0.3807	0.4587	0.8025	0.4306	**0.2255**	0.4618	0.3291
Scene	**0.1689**	0.4617	0.7272	0.6229	0.4215	0.6350	0.5335
Computers	0.6026	0.4153	0.6061	0.4850	**0.2999**	0.4898	0.4782
Average	0.2923	0.3190	0.5358	0.3459	**0.2648**	0.3810	0.2714
Ave. Rank.	3.5	3.6	7	4	**2.2**	5.3	2.4

Table 8
Comparison results of 3CMLL with other algorithms for the RL (↓) metric.

Data set	LLSF-DL	MSWL	GLMAM	GLFS	2SML	CC	3CMLL
CHD49	0.2267	0.2284	0.3915	**0.2133**	0.2271	0.2532	0.2260
Genbase	0.0028	0.0121	0.1690	0.0128	0.4009	**0.0027**	0.0037
Emotions	0.2323	0.2193	0.3754	0.2829	0.2196	0.2843	**0.2190**
Business	0.1112	0.0704	0.2753	0.0511	**0.0368**	0.0886	0.0509
Flags	0.9538	0.2635	0.4475	0.2503	0.2450	0.2673	**0.2306**
Image	0.6633	0.2769	0.4341	0.3138	0.2549	0.3845	**0.1456**
CAL500	0.1845	0.2206	0.4185	0.1850	0.2264	0.2785	**0.1827**
Enron	0.2039	0.2456	0.4647	0.1176	0.2888	0.2910	**0.1112**
Scene	0.9443	0.1959	0.4203	0.2956	**0.1713**	0.2986	0.3380
Computers	0.4096	0.1492	0.3312	**0.1003**	0.2836	0.2104	0.1154
Average	0.2592	0.1908	0.3727	0.1823	0.2354	0.2359	**0.1623**
Ave. Rank.	4.8	3.5	6.5	3	3.5	4.8	**1.9**

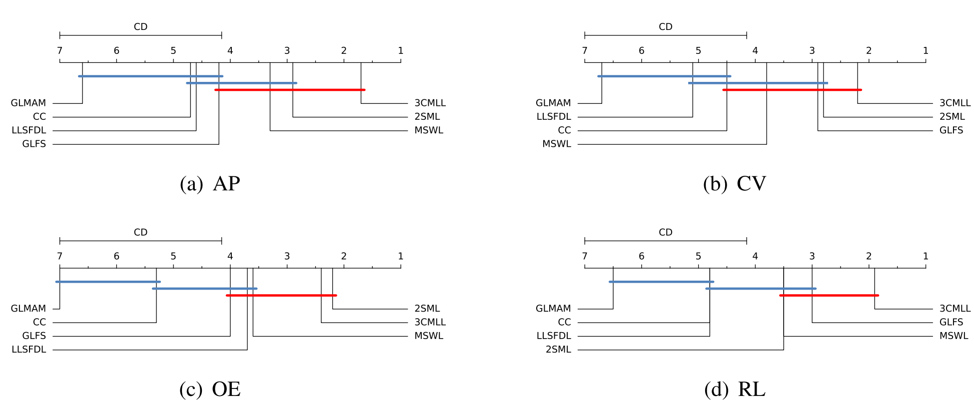
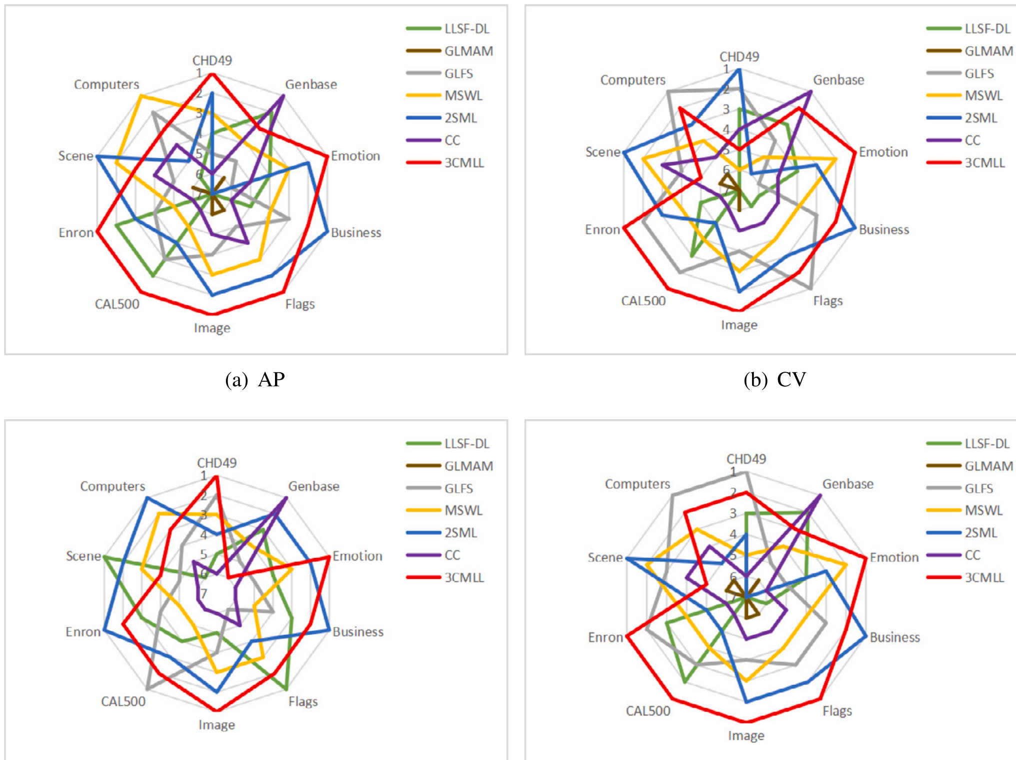


Fig. 5. The Nemenyi test of 3CMLL is compared with other comparison algorithms.



(a) AP (b) CV

(c) OE (d) RL

Fig. 6. Comparison of 3CMLL against other comparison algorithms using radar map.

Table 9
Friedman test and the critical value.

Metric	F_F	Critical value ($\alpha = 0.05$)
AP	12.4772	2.949
CV	12.5589	
OE	18.3253	
RL	10.2857	

dataset. The higher the ranking, the more peripheral the graph of the algorithm is, as depicted in Fig. 6. It is evident from Fig. 6 that the 3CMLL algorithm presents a larger graph in each metric, indicating that 3CMLL is superior to these algorithms.

4.3. Parameter analysis

Within this section, we will conduct sensitivity analysis on the parameters involved in 3CMLL. Except for the parameters to be analyzed, all other parameters are fixed at the optimal parameters. For each data set, the parameter step size to be analyzed is set to 0.1. The influence of parameter β_1, β_2 and α on the evaluation metrics AP, CV, OE and RL is shown in Fig. 7. These experimental results allow for the following conclusions to be obtain:

- On most data sets, β_1 and β_2 achieve high performance on some intermediate values, indicating that overly strict construction of positive and negative key feature sets can lead to performance degradation. It is worth noting that when $\beta_1 \in [0.2, 0.5]$, the optimal values of AP, CV, OE and RL account for $60\%, 60\%, 50\%$ and 60% of all data sets, respectively. Additionally, when $\beta_2 \in [0.5, 0.7]$, the optimal values of AP, CV, OE and RL account for $80\%, 70\%, 80\%$ and 50% of all data sets, respectively.
- The parameter α achieves high performance in some intermediate values on most data sets, indicating that not all concepts generated are worth using. If α is too low, it will be interfered by some invalid concepts, while if α is too high, it will make the concept generalization in the concept space too low, both of which may lead to performance degradation. It is worth noting that when $\alpha \in [0.4, 0.6]$, the optimal values of AP, CV, OE and RL account for $90\%, 80\%, 90\%$ and 80% of all data sets, respectively. Therefore, α in interval $[0.4, 0.6]$ should deserve more attention.

5. Conclusion

This paper proposes a multi-label classification algorithm based on concept-cognitive learning, which considers the correlation between objects, features and labels. Through concept-cognition learning, we obtain feature concepts and multi-label concepts from feature space and label space, respectively. Subsequently, the extent is used as a bridge to correlate the feature concepts with the multi-label concept, and the correlation concept is constructed. Furthermore, effective concepts are extracted from the correlation concept spaces to realize multi-label prediction. On ten multi-label datasets, comprehensive comparisons with the state-of-the-art multi-label classification approaches demonstrate the competitive performance of our approach.

It can be found that the proposed method is based on the regular formal context and cannot directly deal with continuous values. In future

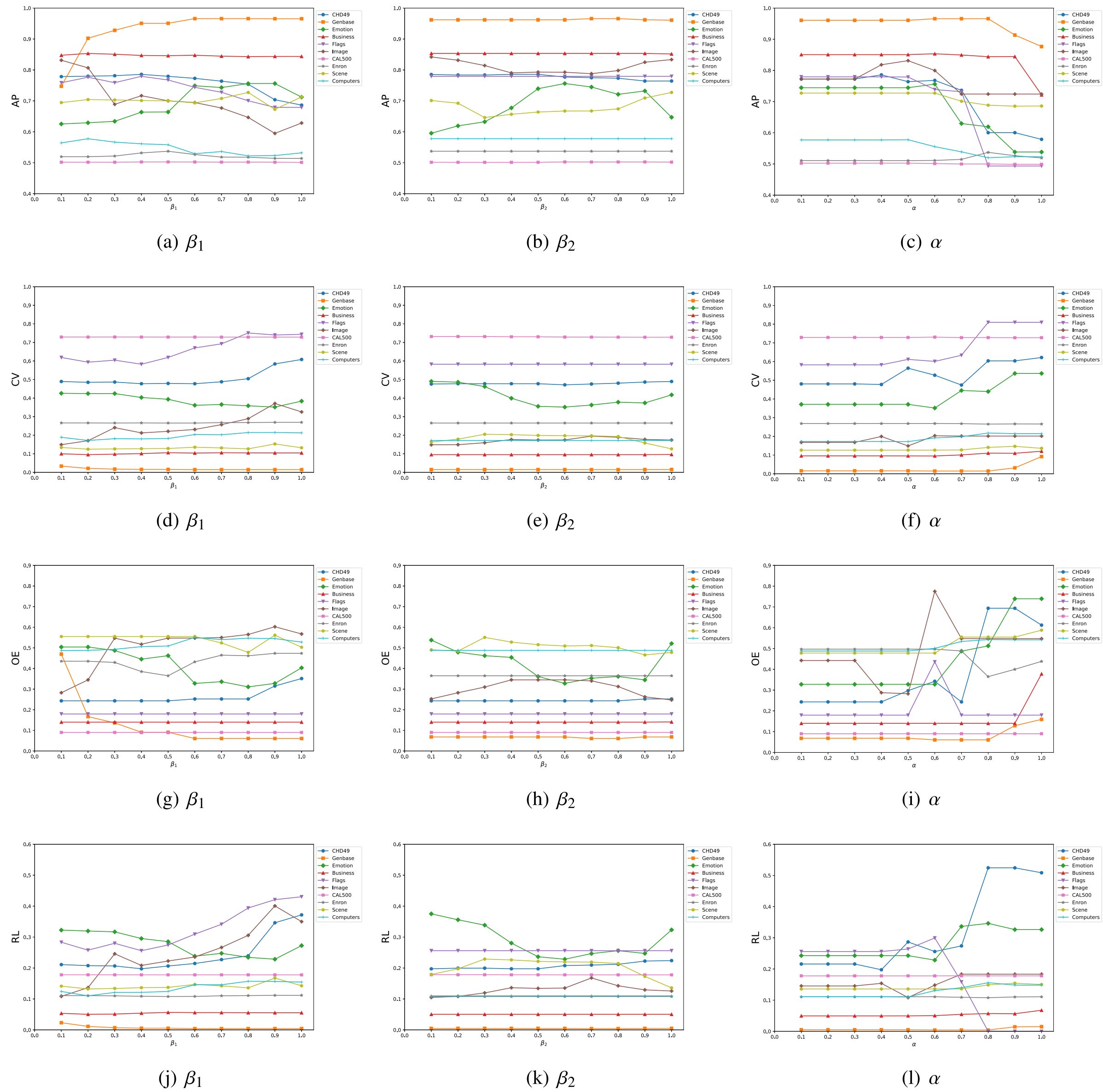


Fig. 7. The influence of parameters β_1, β_2 and α on each metric respectively.

work, we will take into account the multi-label classification algorithm based on concept-cognition in fuzzy formal context. In addition, the discussion of missing labels is also a topic worthy of study.

CRediT authorship contribution statement

Jiaming Wu: Writing – review & editing, Writing – original draft, Software, Methodology, Conceptualization. **Eric C.C. Tsang:** Validation, Supervision, Methodology, Investigation. **Weihua Xu:** Validation, Supervision. **Chengling Zhang:** Visualization, Software. **Lanzhen Yang:** Investigation, Data curation.

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.

Data availability

Data will be made available on request.

Acknowledgments

This paper is supported by the Macao Science and Technology Development Fund (0036/2023/RIA1) and the Natural Science Foundation of China (No. 61976245)

References

[1] J.H. Li, C.L. Mei, W.H. Xu, Y.H. Qian, Concept learning via granular computing: A cognitive viewpoint, Inform. Sci. 298 (2015) 447–467.

[2] W.H. Xu, W.T. Li, Granular computing approach to two-way learning based on formal concept analysis in fuzzy dataset, IEEE Trans. Cybern. 46 (2016) 366–379.

[3] W.H. Xu, D.D. Guo, J.S. Mi, et al., Two-way concept-cognitive learning via concept movement viewpoint, IEEE Trans. Neural. Netw. Learn. (2023) http://dx.doi.org/10.1109/TNNLS.2023.3235800.

[4] J.J. Niu, C.C. Huang, J.H. Li, M. Fan, Parallel computing techniques for concept-cognitive learning based on granular computing, Int. J. Mach. Learn. Cybern. 9 (2018) 1785–1805.

[5] E.C.C. Tsang, B.J. Fan, D.G. Chen, W.H. Xu, W.T. Li, Multi-level cognitive concept learning method oriented to data sets with fuzziness: A perspective from features, Soft. Comput. 24 (2020) 3753–3770.

[6] C.F. Zou, H.F. Deng, J.F. Wan, Z.R. Wang, P. Deng, Mining and updating association rules based on fuzzy concept lattice, Future Gener. Comp. Syst. 82 (2018) 698–706.
[7] J.J. Niu, D.G. Chen, J.H. Li, H. Wang, Fuzzy rule based classification method for incremental rule learning, IEEE Trans. Fuzzy Syst. 30 (9) (2022) 3748–3761.
[8] R. Wille, in: I. Rival (Ed.), Restructuring Lattice Theory: An Approach Based on Hierarchies of Concepts, Ordered Sets, Springer, Dordrecht, The Netherlands, 1982, pp. 445–470.
[9] J.M. Ma, W.X. Zhang, Axiomatic characterizations of dual concept lattices, Internat. J. Approx. Reason. 54 (5) (2013) 690–697.
[10] W.H. Xu, D.D. Guo, Y.H. Qian, W.P. Ding, Two-way concept-cognitive learning method: A fuzzy-based progressive learning, IEEE Trans. Fuzzy Syst. 31 (6) (2023) 1885–1899.
[11] W.X. Zhang, J.M. Ma, S.Q. Fan, Variable threshold concept lattices, Inform. Sci. 177 (2007) 4883–4892.
[12] D.D. Guo, W.H. Xu, Y.H. Qian, W.P. Ding, M-FCCL: Memory-based concept-cognitive learning for dynamic fuzzy data classification and knowledge fusion, Inf. Fusion 100 (2023) 101962.
[13] J.H. Li, C.C. Huang, J.J. Qi, Y.H. Qian, W.Q. Liu, Three-way cognitive concept learning via multi-granularity, Inform. Sci. 378 (2017) 244–263.
[14] D.X. Chen, J.H. Li, R.D. Lin, Formal concept analysis of multi-scale formal context, J. Ambient Intell. Humaniz. Comput. 11 (11) (2020) 5315–5327.
[15] Y.Y. Yao, Concept lattices in rough set theory, in: IEEE Annual Meeting of the Fuzzy Information, NAFIPS, 2004, pp. 796–801.
[16] D.D. Guo, C.M. Jiang, P. Wu, Three-way decision based on confidence level change in rough set, Internat. J. Approx. Reason. 143 (2022) 57–77.
[17] D.D. Guo, C.M. Jiang, R.X. Sheng, S.S. Liu, A novel outcome evaluation model of three-way decision: A change viewpoint, Inform. Sci. 607 (2022) 1089–1110.
[18] Z.H. Huang, J.J. Li, Y.H. Qian, Noise-tolerant fuzzy-β-covering-based multigranulation rough sets and feature subset selection, IEEE Trans. Fuzzy Syst. 30 (7) (2022) 2721–2735.
[19] Z.H. Huang, J.J. Li, C.Z. Wang, Robust feature selection using multigranulation variable-precision distinguishing indicators for fuzzy covering decision systems, IEEE Trans. Syst. Man Cybern. (2023) http://dx.doi.org/10.1109/TSMC.2023.3321315.
[20] Y.Y. Yao, Interpreting concept learning in cognitive informatics and granular computing, IEEE Trans. Syst. Man Cybern. 39 (4) (2009) 855–866.
[21] W.Z. Wu, Y. Leung, J.S. Mi, Granular computing and knowledge reduction in formal contexts, IEEE Trans. Knowl. Data Eng. 21 (10) (2009) 1461–1474.
[22] Y. Shi, Y.L. Mi, J.H. Li, W.Q. Liu, Concurrent concept-cognitive learning model for classification, Inform. Sci. 496 (2019) 65–81.
[23] Y. Shi, Y.L. Mi, J.H. Li, W.Q. Liu, Concept-cognitive learning model for incremental concept learning, IEEE Trans. Syst. Man Cybern. Syst. 51 (2) (2021) 809–821.
[24] Y.L. Mi, Y. Shi, J.H. Li, W.Q. Liu, M.Y. Yan, Fuzzy-based concept learning method: exploiting data with fuzzy conceptual clustering, IEEE Trans. Cybern. 52 (1) (2022) 582–593.
[25] Y.L. Mi, P. Quan, Y. Shi, Z.R. Wang, Concept-cognitive computing system for dynamic classification, European J. Oper. Res. 301 (1) (2022) 287–299.
[26] Y.L. Mi, W.Q. Liu, Y. Shi, et al., Semi-supervised concept learning by concept-cognitive learning and concept space, IEEE Trans. Knowl. Data Eng. 34 (5) (2022) 2429–2442.
[27] K.H. Yuan, W.H. Xu, W.T. Li, W.P. Ding, An incremental learning mechanism for object classification based on progressive fuzzy three-way concept, Inform. Sci. 584 (2022) 127–147.
[28] W.H. Xu, Y.Q. Chen, Multi-attention concept-cognitive learning model: A perspective from conceptual clustering, Knowl.-Based Syst. 252 (2022) 109472.
[29] C.L. Zhang, E.C.C. Tsang, W.H. Xu, Y.D. Lin, L.Z. Yang, Incremental concept-cognitive learning approach for concept classification oriented to weighted fuzzy concepts, Knowl.-Based Syst. 260 (2023) 110093.
[30] M.L. Zhang, Z.H. Zhou, A review on multi-label learning algorithms, IEEE Trans. Knowl. Data Eng. 26 (8) (2014) 1819–1837.
[31] W. Weng, Y.J. Lin, S.X. Wu, Y.W. Li, Y. Kang, Multi-label learning based on label-specific features and local pairwise label correlation, Neurocomputing 273 (2017) 385–394.
[32] J. Zhang, C.D. Li, D.L. Cao, Y.J. Lin, S.Z. Su, L. Dai, S.Z. Li, Multi-label learning with label-specific features by resolving label correlations, Knowl.-Based Syst. 159 (1) (2018) 148–157.
[33] C.M. Zhu, D.Q. Miao, Z. Wang, R.G. Zhou, L. Wei, X.F. Zhang, Global and local multi-view multi-label learning, Neurocomputing 371 (2020) 67–77.
[34] C.M. Zhu, Y.C. Liu, D.Q. Miao, Y.L. Dong, W. Pedrycz, Within- cross- consensus-view representation-based multi-view multi-label learning with incomplete data, Neurocomputing 557 (2023) 126729.
[35] X.Y. Che, D.G. Chen, J.S. Mi, Label correlation in multi-label classification using local attribute reductions with fuzzy rough sets, Fuzzy Sets and Systems 426 (15) (2022) 121–144.
[36] Y.L. Fan, J.H. Liu, P.Z. Liu, Y.Z. Du, W.Y. Lan, S.X. Wu, Manifold learning with structured subspace for multi-label feature selection, Pattern Recognit. 120 (2021) 108169.
[37] M.R. Boutell, J.B. Luo, X.P. Shen, C.M. Brown, Learning multi-label scene classification, Pattern Recognit. 37 (9) (2004) 1757–1771.
[38] M.L. Zhang, L. Wu, LIFT: Multi-label learning with label-specific features, IEEE Trans. Pattern Anal. Mach. Intell. 37 (2015) 107–120.
[39] M.L. Zhang, Z.H. Zhou, ML-KNN: a lazy learning approach to multi-label learning, Pattern Recognit. 40 (7) (2007) 2038–2048.
[40] J. Read, B. Pfahringer, G. Holmes, E. Frank, Classifier chains for multi-label classification, Mach. Learn. 85 (3) (2011) 333–359.
[41] G. Tsoumakas, I. Katakis, I. Vlahavas, Random k-labelsets for multilabel classification, IEEE Trans. Knowl. Data Eng. 23 (7) (2010) 1079–1089.
[42] X.T. Zhen, M.Y. Yu, X.F. He, S. Li, Multi-target regression via robust low-rank learning, IEEE Trans. Pattern Anal. Mach. Intell. 40 (2) (2017) 497–504.
[43] J.L. Li, P.P. Li, X.G. Hu, K.Y., Learning common and label-specific features for multi-label classification with correlation information, Pattern Recognit. 121 (2022) 108259.
[44] Y.L. Fan, J.H. Liu, S.X. Wu, Exploring instance correlations with local discriminant model for multi-label feature selection, Appl. Intell. 52 (2022) 8302–8320.
[45] Y.L. Fan, B.H. Chen, W.Q. Huang, J.H. Liu, W.W., W.Y. Lan, Multi-label feature selection based on label correlations and feature redundancy, Knowl.-Based Syst. 241 (2022) 108256.
[46] J. Huang, G.R. Li, Q.M. Huang, X.D. Wu, Learning label-specific features and class-dependent labels for multi-label classification, IEEE Trans. Knowl. Data Eng. 28 (2016) 3309–3323.
[47] J. Zhang, S.Z. Li, M. Jiang, K.C. Tan, Learning from weakly labeled data based on manifold regularized sparse model, IEEE Trans. Cybern. 52 (5) (2022) 3841–3854.
[48] Y.S. Chen, K. Qian, F. Min, Global and local attention-based multi-label learning with missing labels, Inform. Sci. 594 (2022) 20–42.
[49] J. Zhang, H.R. Wu, M. Jiang, J.H. Liu, S.Z. Li, Y. Tang, J.Y. Long, Group-preserving label-specific feature selection for multi-label learning, Expert Syst. Appl. 213 (2023) 118861.
[50] K. Qian, X.Y. Min, Y.S. Cheng, F. Min, Weight matrix sharing for multi-label learning, Pattern Recognit. 136 (2023) 109156.
[51] J. Demsar, Statistical comparisons of classifiers over multiple data sets, J. Mach. Learn. Res. 7 (2006) 1–30.