LABEL ENHANCEMENT VIA FUZZY CONCEPT ANALYSIS IN MULTI-LABEL LEARNING

JIA-MING WU[1], ERIC C.C. TSANG[1,*], SUO-YU ZHANG[1], YI-MING WANG[1]

[1]School of Computer Science and Engineering, Macau University of Science and Technology, Taipa, Macau, China
E-MAIL: wjmfuzzy@163.com, cctsang@must.edu.mo

Abstract:

Traditional multi-label learning methods often rely on explicit logical labels, making it difficult to uncover the latent label structures hidden within the data, which in limited learning effectiveness. To enhance the representation capacity of sample-label relationships in multi-label learning, this work introduces a concept analysis approach to discover the implicit label distribution information underlying the observable labels, thereby enriching the original label representation. Specifically, concept cognitive learning is employed to model the intrinsic associations between samples and labels, resulting in the construction of a label distribution matrix that captures label distribution characteristics. This matrix, combined with label correlations, is then used to impose constraints on the objective function during the learning process, ultimately improving the performance of multi-label learning.

Keywords:

Multi-label learning; Label distribution learning; Concept-cognitive learning; Label enhancement

1. Introduction

Traditional supervised learning methods typically only allow for assigning a single label to each sample, ignoring the reality that samples may be associated with multiple labels simultaneously. Multi-label learning models the complex dependencies between samples and multiple labels, offering a more accurate representation of the rich and diverse semantic structures found in real-world data [1, 2]. Therefore, multi-label learning has become an important research direction for dealing with complex real-world data, and it has been widely applied in areas such as image annotation, text classification and so on.

In the research of multi-label learning, label enhancement has become an important issue. Its goal is to reconstruct a more detailed and continuous label distribution in the case of only logical labels, so that the label distribution learning method can play a role in data sets lacking clear distribution information [3, 4]. The existing label enhancement methods are mainly divided into two categories: adaptive and specialized algorithms. The adaptive algorithm is an extension and improvement based on the original learning model. A current research hotspot is to introduce fuzzy mathematics into label enhancement, using methods such as fuzzy clustering, fuzzy computation, and fuzzy association analysis to transform discrete logical labels into more expressive label distributions. In contrast, specialized algorithms often adopt graph-based strategies to construct topological structures among samples. These approaches exploit structural assumptions to model latent relationships between samples and labels. By explicitly representing the structural information in feature space, they help uncover deeper semantic associations between labels. Regardless of the approach, effectively converting logical labels into soft label distributions fundamentally depends on exploring the structural properties of the feature space. Analyzing the hierarchical relationships between samples provides essential support for semantic association modeling and label distribution construction.

To improve the quality of label enhancement, this paper introduces Concept-Cognitive Learning (CCL) as a framework for knowledge modeling. CCL is an effective cognitive tool that simulates human learning processes by acquiring concepts and uncovering the hierarchical structures among them [5]. This makes complex knowledge easier to understand and apply. In CCL, each concept consists of two core components: extent and intent, which are mutually determined and form the basic semantic unit. In knowledge representation, the involution relationship between the intent and extent of the concept effectively reveals the internal structure of knowledge [6, 7]. In recent years, benefiting from the advantages of CCL in structural modeling and semantic expression, researchers have attempted to introduce it into multi-label learning and developed various concept-cognitive learning models [8, 9]. However, most of the existing methods still remain in the modeling of explicit logical labels, and the potential semantic information of labels has not been fully explored. To address this gap, we propose a CCL-based label enhancement method. By leveraging concept analysis to extract hierarchical structure from the feature space, the proposed method facilitates the transformation of logical labels into continuous numerical representations. This not only enhances the expressiveness of labels but also improves the overall performance of multi-label learning models.

The remainder of this paper is organized as follows. In Section 2, we briefly review label distribution learning and fuzzy formal concept analysis. In Section 3, we describe the proposed method. Comparative experimental results and analyses are presented in Section 4. Finally, Section 5 provides the conclusion.

2. Preliminaries

2.1. Label distribution learning

In multi-label learning, each instance can be associated with multiple semantic labels simultaneously. Let the multi-label dataset be denoted as $\mathcal{D} = (\mathbf{x}_i, \mathbf{y}_i)$, $i \in \{1, 2, \cdots, n\}$, where $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_n]^T \in \mathbb{R}^{n\times m}$ represents the feature matrix of n instances, and each instance $\mathbf{x}_i = [x_{i1}, x_{i2}, \cdots, x_{im}]$ is an m-dimensional feature vector. The corresponding label matrix is $\mathbf{Y} = [\mathbf{y}_1, \mathbf{y}_2, \cdots, \mathbf{y}_n]^T \in \{0, 1\}^{n\times q}$, where each label vector $\mathbf{y}_i = [y_{i1}, y_{i2}, \cdots, y_{iq}]$ indicates the relevance of q possible labels to the i-th instance. Specifically, $y_{ij} = 1$ denotes that the j-th label is relevant to the i-th instance; $y_{ij} = 0$ denotes that the j-th label is not relevant to the i-th instance.

In label distribution learning, each possible label y_{ij} for a given sample x_i is assigned a continuous value, denoted as d_{ij}, which indicates the degree to which the label describes the sample. This label distribution must satisfy the following two conditions:
(1) $d_{ij} \in [0, 1]$, meaning that the description degree for each label lies within the interval [0, 1];
(2) $\sum_j^q d_{ij} = 1$, ensuring that the total description degree across all labels for a given sample sums to 1.

This formulation allows each sample to be characterized by a complete label distribution vector rather than a set of discrete logical labels, thereby capturing the varying degrees of relevance between the sample and each label more precisely.

2.2. Fuzzy formal concept analysis

A fuzzy formal context is defined as a triplet $\left(U, F, \tilde{I}\right)$, where:

- $U = \{o_1, o_{2,} \cdots, o_n\}$ is the set of objects.
- $F = \{f_1, f_{2,} \cdots, f_m\}$ is the set of attributes.
- $\tilde{I}$ is a fuzzy relation between U and F, such that $\tilde{I} : U \times F \to [0, 1]$.

The fuzzy relationship $\tilde{I}(o, f)$ represents the degree to which an object $o \in U$ possesses the feature $f \in F$. Let $\mathcal{P}(U)$ denote the power set of U, and $\mathcal{F}(F)$ denote the set of all fuzzy subsets on F. For any $f \in F$, the membership function $\tilde{F}(f) : F \to [0, 1]$ indicates the degree to which the feature f belongs to a given fuzzy subset $\tilde{F} \in \mathcal{F}(F)$. For $O \in \mathcal{P}(U)$, $\tilde{A} \in \mathcal{F}(F)$, the operators $\uparrow$: $\mathcal{P}(U) \to \mathcal{F}(F)$ and $\downarrow$: $\mathcal{F}(F) \to \mathcal{P}(U)$ are defined as:

$$O^{\uparrow}(f) = \bigwedge_{o\in O} \tilde{I}(o, f), f \in F, \tag{1}$$

$$\tilde{A}^{\downarrow} = \left\{ o \in U \,\middle|\, \forall f \in F, \tilde{A}(f) \leqslant \tilde{I}(o, f) \right\}. \tag{2}$$

If $O^{\uparrow} = \tilde{A}$ and $\tilde{A}^{\downarrow} = O$, then the pair $\left(O, \tilde{A}\right)$is called a fuzzy concept, where O and $\tilde{A}$ are called the extent and intent, respectively.

3. Multi-label classification with weakly labeled data

A multi-label context can be represented by the triple $\langle U, F, \mathcal{D}\rangle$, where $U = \{o_1, o_2, \cdots, o_n\}$ is the set of objects, $F = \{f_1, f_2, \cdots, f_m\}$ is the set of attributes, and $\mathcal{D} = (\mathbf{x}_i, \mathbf{y}_i)$ is the multi-label dataset. Let $\mathcal{P}(U)$ denote the power set of U. For any $f \in F$, the function $\tilde{F}(f) : F \to [0, 1]$ represents the membership degree of the attribute f in the fuzzy set $\tilde{F}$, where $\tilde{F} = \left\{\tilde{F}(f_1), \tilde{F}(f_2), \cdots, \tilde{F}(f_m)\right\}$. The set of all fuzzy subsets over F is denoted by $\mathcal{F}(F)$.

Given a multi-label context $\langle U, F, \mathcal{D} \rangle$, for any $O \in \mathcal{P}(U)$ and $\tilde{A} \in \mathcal{F}(F)$, we define the operators $\mathcal{L} : \mathcal{P}(U) \to \mathcal{F}(F)$ and $\mathcal{G} : \mathcal{F}(F) \to \mathcal{P}(U)$ as follows:

$$\mathcal{L}(O)(f_j) = \bigwedge_{o_i \in O} x_{ij}, f_j \in F,$$
$$\mathcal{G}\left(\tilde{A}\right) = \left\{ o_i \in U \middle| \forall f_j \in F, \tilde{A}(f_j) \leqslant x_{ij} \right\}.$$

A pair $\left(O, \tilde{A}\right)$ is called a fuzzy concept, if it satisfies $\mathcal{G}\left(\tilde{A}\right) = O$ and $\mathcal{L}(O) = \tilde{A}$, where O and $\tilde{A}$ are referred to as the extent and intent of the concept, respectively. Clearly, the pair $(\mathcal{GL}(O), \mathcal{L}(O))$ constitutes a fuzzy concept induced by the object set O.

For a given label index $j \in \{1, 2, \cdots, q\}$, let $O_j^+ = \{o_i \in U | y_{ij} = 1\}$ and $O_j^- = \{o_i \in U | y_{ij} = 0\}$ denote the sets of objects that are respectively relevant and irrelevant to the j-th label. It follows that the object set U can be partitioned with respect to label j as $O_j^+ \cup O_j^- = U$ and $O_j^+ \cap O_j^- = \emptyset$. This partition forms the basis for analyzing label-specific relevance in a multi-label learning context.

Let $\langle U, F, \mathcal{D} \rangle$ be a multi-label context. For a given label l_j and its associated object sets $O_j^+ = \{o_i \in U | y_{ij} = 1\}$ and $O_j^- = \{o_i \in U | y_{ij} = 0\}$, the positive concept space and negative concept space of label l_j are defined as follows:

$$\mathcal{S}_j^+ = \left\{ (\mathcal{GL}(o), \mathcal{L}(o)) \middle| o \in O_j^+ \right\},$$
$$\mathcal{S}_j^- = \left\{ (\mathcal{GL}(o), \mathcal{L}(o)) \middle| o \in O_j^- \right\}.$$

Here, each pair $(\mathcal{GL}(o), \mathcal{L}(o))$ represents a fuzzy concept induced by object $o \in U$. The set $\mathcal{S}_j^+$ collects all such fuzzy concepts corresponding to objects positively associated with label l_j, while $\mathcal{S}_j^-$ collects those corresponding to negatively associated objects.

Given a multi-label context $\langle U, F, \mathcal{D} \rangle$, let $\mathcal{S}_j^+$ and $\mathcal{S}_j^-$ denote the positive and negative concept spaces of label l_j, where each element is a fuzzy concept in the form $\left(X_k^+, \tilde{B}_k^+\right) \in \mathcal{S}_j^+$ and $\left(X_k^-, \tilde{B}_k^-\right) \in \mathcal{S}_j^-$. The overall conceptual cognition for label l_j is defined as two representative fuzzy concepts:
The positive cognition of label l_j:

$$pc_j^+ = \left(E_j^+, \tilde{I}_j^+\right);$$

The negative cognition of label l_j:

$$pc_j^- = \left(E_j^-, \tilde{I}_j^-\right).$$

These are constructed based on the most representative fuzzy concepts in $\mathcal{S}_j^+$ and $\mathcal{S}_j^-$.
For the positive part:
Let

$$K_j^+ = arg \max_{k \in \left\{1, 2, \cdots, \left|\mathcal{S}_j^+\right|\right\}} \left|X_k^+\right|,$$

then define

$$\left(E_j^+, \tilde{I}_j^+\right) = \begin{cases} \left(X_k^+, \tilde{B}_k^+\right) & , if\ K_j^+ = \{k\}; \\ \left(\bigcup_{k \in K_j^+} X_k^+, \frac{1}{\left|K_j^+\right|} \sum_{k \in K_j^+} \tilde{B}_k^+\right) & , if\ \left|K_j^+\right| > 1. \end{cases}$$

For the negative part:
Similarly, let

$$K_j^- = arg \max_{k \in \left\{1, 2, \cdots, \left|\mathcal{S}_j^-\right|\right\}} \left|X_k^-\right|,$$

then define

$$\left(E_j^-, \tilde{I}_j^-\right) = \begin{cases} \left(X_k^-, \tilde{B}_k^-\right) & , if\ K_j^- = \{k\}; \\ \left(\bigcup_{k \in K_j^-} X_k^-, \frac{1}{\left|K_j^-\right|} \sum_{k \in K_j^-} \tilde{B}_k^-\right) & , if\ \left|K_j^-\right| > 1. \end{cases}$$

Here, E_j^+ and E_j^- represent the aggregated object sets (extents), while $\tilde{I}_j^+$ and $\tilde{I}_j^-$ are the corresponding averaged fuzzy attribute sets (intents). Then, the set of overall concepts by all labels is $\mathcal{PC} = \left\{ \left(pc_1^+, pc_1^-\right), \left(pc_2^+, pc_2^-\right), \cdots, \left(pc_q^+, pc_q^-\right) \right\}$.

Given a multi-label context $\langle U, F, \mathcal{D} \rangle$, and the overall conceptual cognition for each label l_j denoted by $\left(E_j^+, \tilde{I}_j^+\right)$ and $\left(E_j^-, \tilde{I}_j^-\right)$, we define the distance-based label relevance for each instance as:

$$d_{ij} = \begin{cases} \left\| \mathbf{x}_i - \tilde{I}_j^+ \right\|_2 & , if\ y_{ij} = 1; \\ -\left\| \mathbf{x}_i - \tilde{I}_j^- \right\|_2 & , if\ y_{ij} = 0. \end{cases}$$

Let $\mathbf{d}_i = \{d_{i1}, d_{i2}, \cdots, d_{iq}\}$ denote the collection of distances between instance $\mathbf{x}_i$ and all concepts. To ensure consistency and comparability across labels, these distances are normalized as:

$$d_{ij}^{norm} = \frac{d_{ij} - \min(\mathbf{d}_i)}{\max(\mathbf{d}_i) - \min(\mathbf{d}_i)},$$
$$\hat{d}_{ij} = \frac{d_{ij}^{norm}}{\sum_{k=1}^{q} d_{ik}^{norm}}.$$

The label distribution matrix used for label enhancement is then defined as:

$$\mathbf{D} = \left(\hat{d}_{ij}\right)_{n \times q}$$

where each entry $\hat{d}_{ij}$ represents the normalized semantic distance between instance $\mathbf{x}_i$ and the corresponding positive or negative fuzzy concept of label l_j. This matrix $\mathbf{D}$ captures the relative label distances and can serve as a soft label distribution for further label enhancement tasks. Based on the above basis, algorithm 1 is given.

4. Experiments

4.1. Datasets and experiment settings

To further evaluate the effectiveness of the proposed algorithm, experiments were conducted on 10 publicly available multi-label datasets. The detailed characteristics of these datasets are summarized in Table 1. Specifically, n, m, and q denote the number of instances, features, and labels, respectively. Additionally, "Card" represents the average label cardinality, while "Domain" indicates the type of dataset.

To assess the predictive performance of the proposed algorithm, it is compared against three representative multi-label learning methods (ML-KNN [10], MDFS [11] and MRDM [12]).

4.2. Evaluation metrics

Evaluating performance in multi-label classification is inherently more challenging than in traditional single-label settings, due to the presence of multiple, potentially correlated labels. To effectively assess algorithm performance in this context, four widely used evaluation metrics are employed [1]: Average Precision (AP), Coverage (CV), One-error (OE), and Ranking Loss (RL). These metrics collectively offer a comprehensive assessment of how well an algorithm captures the complex characteristics of multi-label data. Given a test set $\mathcal{T} = (x_i, L_i) \mid 1 \leq i \leq s$, where L_i denotes the set of relevant labels for instance x_i, and prediction scores for all labels are sorted in descending order as $f_1(x_i), f_2(x_i), \ldots, f_q(x_i)$, the definitions of the four evaluation metrics are presented below:

Average Precision (AP): This metric evaluates the average likelihood that relevant labels are ranked above irrelevant ones. A higher average precision indicates that the classifier is more effective at prioritizing relevant labels in its ranking.

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

Algorithm 1: The label distribution matrix.

Input: A multi-label context $\langle U, F, \mathcal{D} \rangle$.
Output: The label distribution matrix $\mathbf{D}$.

$\mathcal{PC} = \emptyset$;
for $j = 1 : q$ do
    $O_j^+ = \emptyset$; $O_j^- = \emptyset$; $\mathcal{S}_j^+ = \emptyset$; $\mathcal{S}_j^- = \emptyset$;
    for $i = 1 : n$ do
        if $y_{ij} = 1$ then
            $O_j^+ \leftarrow o_i$;
            Compute fuzzy concept $\left(\mathcal{GL}\left(o_i\right), \mathcal{L}\left(o_i\right)\right)$;
            $\mathcal{S}_j^+ \leftarrow \left(\mathcal{GL}\left(o_i\right), \mathcal{L}\left(o_i\right)\right)$;
        else
            $O_j^- \leftarrow o_i$;
            Compute fuzzy concept $\left(\mathcal{GL}\left(o_i\right), \mathcal{L}\left(o_i\right)\right)$;
            $\mathcal{S}_j^- \leftarrow \left(\mathcal{GL}\left(o_i\right), \mathcal{L}\left(o_i\right)\right)$;
        end
    end
    Calculate K_j^+ and K_j^-;
    if $\left|K_j^+\right| = 1$ then
        $E_j^+ = X_k^+$;
        $\tilde{I}_j^+ = \tilde{B}_k^+$;
    else
        $E_j^+ = \bigcup_{k \in K_j^+} X_k^+$;
        $\tilde{I}_j^+ = \frac{1}{\left|K_j^+\right|} \sum_{k \in K_j^+} \tilde{B}_k^+$;
    end
    $pc_j^+ = \left(E_j^+, \tilde{I}_j^+\right)$; $pc_j^- = \left(E_j^-, \tilde{I}_j^-\right)$;
    $\mathcal{PC} \leftarrow \left(pc_j^+, pc_j^-\right)$;
end
for $i = 1 : n$ do
    for $j = 1 : q$ do
        if $y_{ij} = 1$ then
            $d_{ij} = \left\|\mathbf{x}_i - \tilde{I}_j^+\right\|_2$;
        else
            $d_{ij} = -\left\|\mathbf{x}_i - \tilde{I}_j^-\right\|_2$;
        end
    end
end
Calculate $\hat{d}_{ij}$;
$\mathbf{D} = \left(\hat{d}_{ij}\right)_{n \times q}$;
return $\mathbf{D}$.

TABLE 1. Description of the multi-label datasets.

Dataset	n	m	q	Card	Domain
Emotion	593	72	6	1.869	Music
Birds	645	260	19	1.014	Audio
Enron	1702	1001	53	3.378	Text
Image	2000	294	5	1.236	Image
Genbase	662	1186	27	1.252	Biology
Flags	194	19	7	3.392	Image
CAL500	502	68	174	26.044	Music
CHD49	555	49	6	2.580	Medicine
Medical	978	1449	45	1.245	Text
Scene	2407	294	6	1.074	Image

where, $R_i = \{l_j | rank(x_i, l_j) \leqslant rank(x_i, l_k)\}$.

Coverage (CV): This metric indicates the number of positions one must scan down the predicted label ranking to include all relevant labels. A lower coverage value means that all true labels appear closer to the top of the ranking, reflecting better classifier performance.

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

One-error (OE): This metric measures the fraction of instances where the top-ranked predicted label is not among the true relevant labels. A lower one-error value indicates that the classifier is more accurate in identifying the most relevant label.

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

Ranking Loss (RL): A lower ranking loss indicates better performance of the multi-label classification model, as it reflects fewer instances where irrelevant labels are ranked above relevant ones.

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

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

4.3. Experimental results

In this section, we evaluate the proposed method using five-fold cross-validation. Each dataset is randomly partitioned into five equal subsets. During each iteration, one subset is used for testing while the remaining four are used for training. This process is repeated five times, ensuring each subset serves as the test set once. The final performance is reported as the average over the five runs. We assess the results using four commonly used evaluation metrics: Average Precision (↑), Coverage (↓), One-error (↓), and Ranking Loss (↓). Here, (↑) indicates that higher values are better, while (↓) indicates that lower values are preferable. All results are presented in the format "mean ± standard deviation", and the best result in each row is highlighted in bold. The complete evaluation outcomes are summarized in Table 2.

As shown in Table 2, the proposed LECCL algorithm achieves leading performance on the majority of datasets across all four evaluation metrics: Average Precision, Coverage, One-error, and Ranking Loss. These results highlight the effectiveness of the proposed label enhancement strategy. By leveraging concept-cognitive learning to reconstruct and enrich label representations, LECCL is able to capture latent label distributions and structural relationships more accurately. This enhanced label information significantly contributes to improved classification performance, demonstrating the advantage of incorporating label enhancement in multi-label learning.

5. Conclusions

This paper proposes a label enhancement method for multi-label learning based on concept-cognitive learning. By simulating the human cognitive process, the method explores the latent structural information and intrinsic correlations among labels within multi-label data to enhance the expressive power of the original labels. This enhanced representation enables the learning model to better capture and interpret complex semantic relationships. Experimental results on ten real-world multi-label datasets demonstrate that the proposed method achieves superior performance across multiple evaluation metrics, validating its effectiveness and advantages in addressing multi-label learning tasks.

Acknowledgements

This paper is supported by the Macao Science and Technology Development Fund (0036/2023/RIA1, 0037/2024/RIA1) and Jiangsu Kaite Auto Parts Co., Ltd. (MUST Code: 1330).

TABLE 2. Comparison of methods based on multiple metrics.

Dataset	Methods	Average Precision(↑)	Coverage(↓)	One-error(↓)	Ranking Loss(↓)
Emotion	LECCL	$0.8035_{\pm 0.0167}$	$0.3002_{\pm 0.0164}$	$0.2565_{\pm 0.0246}$	$0.1596_{\pm 0.0160}$
	ML-KNN	$0.7889_{\pm 0.0187}$	$0.3067_{\pm 0.0205}$	$0.2919_{\pm 0.0329}$	$0.1718_{\pm 0.0174}$
	MDFS	$0.7998_{\pm 0.0259}$	$0.3255_{\pm 0.0169}$	$0.2635_{\pm 0.0463}$	$0.1758_{\pm 0.0219}$
	MRDM	$0.8011_{\pm 0.0046}$	$0.2987_{\pm 0.0177}$	$0.2648_{\pm 0.0038}$	$0.1643_{\pm 0.0065}$
Birds	LECCL	$0.7607_{\pm 0.0214}$	$0.1464_{\pm 0.0165}$	$0.2884_{\pm 0.0192}$	$0.0962_{\pm 0.0120}$
	ML-KNN	$0.7168_{\pm 0.0261}$	$0.1528_{\pm 0.0199}$	$0.3550_{\pm 0.0445}$	$0.1059_{\pm 0.0137}$
	MDFS	$0.7074_{\pm 0.0181}$	$0.1498_{\pm 0.0105}$	$0.3687_{\pm 0.0301}$	$0.1085_{\pm 0.0102}$
	MRDM	$0.7210_{\pm 0.0187}$	$0.1571_{\pm 0.0189}$	$0.3457_{\pm 0.0305}$	$0.1121_{\pm 0.0128}$
Enron	LECCL	$0.7047_{\pm 0.0084}$	$0.2598_{\pm 0.0155}$	$0.2197_{\pm 0.0125}$	$0.0914_{\pm 0.0054}$
	ML-KNN	$0.6239_{\pm 0.0082}$	$0.2578_{\pm 0.0147}$	$0.3137_{\pm 0.0161}$	$0.0938_{\pm 0.0059}$
	MDFS	$0.6253_{\pm 0.0202}$	$0.2566_{\pm 0.0060}$	$0.3187_{\pm 0.0328}$	$0.0926_{\pm 0.0036}$
	MRDM	$0.6341_{\pm 0.0078}$	$0.2427_{\pm 0.0053}$	$0.3184_{\pm 0.0335}$	$0.0892_{\pm 0.0035}$
Image	LECCL	$0.7782_{\pm 0.0102}$	$0.1992_{\pm 0.0097}$	$0.3430_{\pm 0.0135}$	$0.1812_{\pm 0.0119}$
	ML-KNN	$0.7867_{\pm 0.0112}$	$0.1971_{\pm 0.0129}$	$0.3270_{\pm 0.0180}$	$0.1795_{\pm 0.0113}$
	MDFS	$0.7816_{\pm 0.0179}$	$0.2063_{\pm 0.0122}$	$0.3242_{\pm 0.0290}$	$0.1915_{\pm 0.0155}$
	MRDM	$0.7512_{\pm 0.0080}$	$0.2182_{\pm 0.0084}$	$0.3900_{\pm 0.0157}$	$0.2067_{\pm 0.0099}$
Genbase	LECCL	$0.9952_{\pm 0.0038}$	$0.0146_{\pm 0.0056}$	$0.0015_{\pm 0.0030}$	$0.0027_{\pm 0.0030}$
	ML-KNN	$0.3189_{\pm 0.0330}$	$0.2115_{\pm 0.0195}$	$0.8906_{\pm 0.0749}$	$0.1958_{\pm 0.0185}$
	MDFS	$0.9844_{\pm 0.0046}$	$0.0104_{\pm 0.0016}$	$0.0284_{\pm 0.0087}$	$0.0014_{\pm 0.0011}$
	MRDM	$0.9868_{\pm 0.0043}$	$0.0201_{\pm 0.0063}$	$0.0136_{\pm 0.0056}$	$0.0060_{\pm 0.0035}$
Flags	LECCL	$0.8104_{\pm 0.0129}$	$0.5520_{\pm 0.0360}$	$0.2032_{\pm 0.0206}$	$0.2170_{\pm 0.0165}$
	ML-KNN	$0.8033_{\pm 0.0251}$	$0.5491_{\pm 0.0228}$	$0.2293_{\pm 0.0869}$	$0.2256_{\pm 0.0320}$
	MDFS	$0.8020_{\pm 0.0127}$	$0.5718_{\pm 0.0141}$	$0.2207_{\pm 0.0457}$	$0.2350_{\pm 0.0107}$
	MRDM	$0.8058_{\pm 0.0200}$	$0.5472_{\pm 0.0151}$	$0.2293_{\pm 0.0388}$	$0.2252_{\pm 0.0142}$
CAL500	LECCL	$0.4994_{\pm 0.0090}$	$0.7446_{\pm 0.0150}$	$0.1134_{\pm 0.0305}$	$0.1804_{\pm 0.0050}$
	ML-KNN	$0.4937_{\pm 0.0088}$	$0.7513_{\pm 0.0204}$	$0.1183_{\pm 0.0297}$	$0.1845_{\pm 0.0068}$
	MDFS	$0.4959_{\pm 0.0021}$	$0.7499_{\pm 0.0021}$	$0.1002_{\pm 0.0012}$	$0.1857_{\pm 0.0008}$
	MRDM	$0.4899_{\pm 0.0125}$	$0.7504_{\pm 0.0184}$	$0.1176_{\pm 0.0325}$	$0.1846_{\pm 0.0025}$
CHD49	LECCL	$0.7936_{\pm 0.0179}$	$0.4453_{\pm 0.0105}$	$0.2559_{\pm 0.0429}$	$0.2018_{\pm 0.0143}$
	ML-KNN	$0.7715_{\pm 0.0159}$	$0.4817_{\pm 0.0201}$	$0.2342_{\pm 0.0332}$	$0.2310_{\pm 0.0153}$
	MDFS	$0.7655_{\pm 0.0074}$	$0.4844_{\pm 0.0076}$	$0.2424_{\pm 0.0136}$	$0.2534_{\pm 0.0088}$
	MRDM	$0.7938_{\pm 0.0127}$	$0.4522_{\pm 0.0165}$	$0.2354_{\pm 0.0205}$	$0.2089_{\pm 0.0136}$
Medical	LECCL	$0.9071_{\pm 0.0217}$	$0.0255_{\pm 0.0062}$	$0.1268_{\pm 0.0284}$	$0.0157_{\pm 0.0053}$
	ML-KNN	$0.8024_{\pm 0.0100}$	$0.0610_{\pm 0.0067}$	$0.2587_{\pm 0.0121}$	$0.0431_{\pm 0.0070}$
	MDFS	$0.7881_{\pm 0.0185}$	$0.0632_{\pm 0.0073}$	$0.2756_{\pm 0.0214}$	$0.0453_{\pm 0.0064}$
	MRDM	$0.8618_{\pm 0.0104}$	$0.0498_{\pm 0.0076}$	$0.1779_{\pm 0.0151}$	$0.0335_{\pm 0.0070}$
Scene	LECCL	$0.8461_{\pm 0.0102}$	$0.0890_{\pm 0.0087}$	$0.2568_{\pm 0.0137}$	$0.0898_{\pm 0.0090}$
	ML-KNN	$0.8579_{\pm 0.0130}$	$0.0885_{\pm 0.0032}$	$0.2506_{\pm 0.0227}$	$0.0900_{\pm 0.0075}$
	MDFS	$0.8486_{\pm 0.0212}$	$0.0883_{\pm 0.0136}$	$0.2528_{\pm 0.0319}$	$0.0893_{\pm 0.0163}$
	MRDM	$0.7544_{\pm 0.0276}$	$0.1451_{\pm 0.0180}$	$0.3976_{\pm 0.0388}$	$0.1567_{\pm 0.0224}$

References

[1] M.L. Zhang, Z.H. Zhou, "A review on multi-label learning algorithms," IEEE Transactions on Knowledge and Data Engineering, vol. 26, no. 8, pp. 1819–1837, 2014.

[2] W.W. Liu, H.B. Wang, X.B. Shen and I.W. Tsang, "The emerging trends of multi-label learning," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 44, no. 11, pp. 7955–7974, 2022.

[3] X. Geng, "Label distribution learning," IEEE Transactions on Knowledge and Data Engineering, vol. 28, no. 7, pp. 1734–1748, 2016.

[4] N. Xu, Y. Liu, X. Geng, "Label enhancement for label distribution learning," IEEE Transactions on Knowledge and Data Engineering, vol. 33, no. 4, pp. 1632–1643, 2019.

[5] D.D. Guo, W.H. Xu, W.P. Ding, et al., "Concept-cognitive learning survey: mining and fusing knowledge from data," Information Fusion, vol. 109, pp. 102426, 2024.

[6] W.X. Zhang, J.M. Ma, S.Q. Fan, "Variable threshold concept lattices," Information Science, vol. 177, pp. 4883–4892, 2007.

[7] Y.Y. Yao, "Interpreting Concept Learning in Cognitive Informatics and Granular Computing," IEEE Transactions on Systems, Man, and Cybernetics, Part B, vol. 39, no. 4, pp. 855–866, 2009.

[8] J.M. Wu, E. C.C. Tsang, W.H. Xu, et al. , "Correlation concept-cognitive learning model for multi-label classification," Knowledge-Based Systems, vol. 290, pp. 111566, 2024.

[9] J.M. Wu, E. C.C. Tsang, W.H. Xu, et al. , "Multi-level correlation information fusion via three-way concept-cognitive learning for multi-label learning," Information Fusion, vol. 124, pp. 103361, 2025.

[10] M.L. Zhang, Z.H. Zhou, "ML-KNN: a lazy learning approach to multi-label learning," Pattern Recognition, vol. 40, pp. 2038–2048, 2007.

[11] J. Zhang, Z.M. Luo, C.D. Li, et al., "Manifold regularized discriminative feature selection for multi-label learning," Pattern Recognition, vol. 95, pp. 136–150, 2019.

[12] R. Huang, Z.J. Wu, "Multi-label feature selection via manifold regularization and dependence maximization," Pattern Recognition, vol. 120, pp. 108149, 2021.