Label Confidence Recovery with High-order Label Correlation in Partial Multi-label Learning

Yuanchao Dai[1,2], **Ximing Li**[1,2,3*], **Ming-Kun Xie**[3], **Xurui Li**[1,2], **Changchun Li**[1,2]
[1]College of Computer Science and Technology, Jilin University, China
[2]Key Laboratory of Symbolic Computation and Knowledge Engineering, Jilin University, China
[3]RIKEN Center for Advanced Intelligence Project, Japan
{yuanchaodai, liximing86, llgglxr, changchunli93}@gmail.com, ming-kun.xie@riken.jp

Abstract

Partial multi-label learning (PML) addresses weakly-supervised scenarios where each instance is associated with a candidate label set containing both ground-truth and noisy labels. Existing PML methods primarily focus on instance-level features or pairwise label correlations for disambiguation. Building on recent insights that exploiting pairwise label correlations improves disambiguation, we observe that high-order label correlations provide even stronger disambiguation evidence in partial settings because ground-truth labels with high-order correlations frequently co-occur, while noise labels produce inconsistent combinations that rarely repeat. To exploit this property, we propose REHC-PML (Label Confidence **RE**covery with **H**igh-order Label **C**orrelations in **P**artial **M**ulti-label **L**earning), which mines frequent high-order co-occurrences from candidate sets to identify global high-order label correlations, selects instance-relevant correlations via Gumbel-Softmax pruning, and propagates their evidence to constituent labels for confidence recovery. Self-training iteratively refines pseudo-labels and trains the classifier. Extensive experiments on both real-world and UCI datasets demonstrate the effectiveness of REHC-PML.

1 Introduction

Multi-label learning addresses the tasks where each training instance can be associated with multiple class labels [Zhang and Zhou, 2014; Zhang and Wu, 2015]. Unfortunately, in many real-world scenarios, it is extremely challenging to annotate precise supervision due to various reasons such as limitations in annotator expertise or the uncertainty of crowdsourcing. Among them, one common scenario is that each instance is annotated with a candidate label set containing relevant labels but also irrelevant ones. To handle this, the emergent task, namely partial multi-label learning (PML), has recently attracted much more attention in the community [Xie and Huang, 2018; Fang and Zhang, 2019; Zhang and Fang, 2021].

*Corresponding author

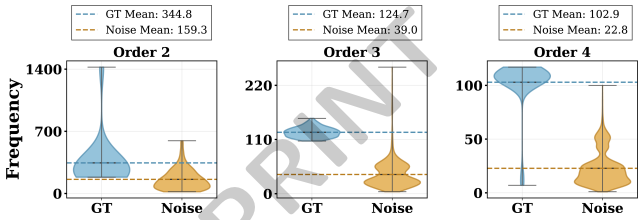

Figure 1: Co-occurrence frequency of ground-truth (GT) *v.s.* noise label combinations across different orders.

Recently, many PML methods have been proposed. One basic problem of PML is how to disambiguate candidate labels while training a multi-label classifier. Roughly, existing methods can be broadly categorized by their disambiguation strategies. Some methods treat each label independently and perform disambiguation by estimating label confidence from candidate annotations or exploiting structural information in the feature space [Zhang and Fang, 2021; Xie *et al.*, 2021; Sun *et al.*, 2019; Yu *et al.*, 2018]. For example, PARTICLE [Zhang and Fang, 2021] estimates labeling confidence via iterative label propagation, and PML-MD [Xie *et al.*, 2021] adaptively learns confidence weights through meta-learning. More recently, many methods recognize that label correlations are inherent in multi-label data and can provide valuable structural signals for disambiguation. For example, GLC [Sun *et al.*, 2022] captures pairwise label correlations through low-rank representation and manifold regularization, while PLAIN [Wang *et al.*, 2023] constructs a label correlation graph to propagate information across label pairs. These pairwise correlation-based methods have demonstrated that exploiting label correlations is valuable for PML disambiguation.

Labels in multi-label data often exhibit joint dependencies among three or more labels instead of pairwise relationships, which are referred to as high-order label correlations [Read *et al.*, 2011; Zhang and Zhang, 2010]. In real-world scenarios, labels frequently co-occur in semantic groups, such as *beach*, *ocean*, and *sunset* in image annotation, and such collective patterns are difficult to capture through pairwise modeling alone. Moreover, high-order correlations offer distinct advantages for disambiguation in partial settings. Since noise labels are introduced randomly, the likelihood of them forming consistent high-order correlations diminishes rapidly as the order increases, while ground-truth label correlations main-

tain stable co-occurrence patterns, resulting in an increasingly pronounced gap between them. As illustrated in Figure 1, the relative frequency gap between ground-truth and noise combinations becomes more pronounced as the order increases, confirming that high-order patterns provide stronger disambiguation signals than pairwise ones.

In this paper, to exploit this property, we propose REHC-PML, a novel PML framework that leverages high-order label correlations for effective disambiguation. The framework operates in two phases. First, we mine frequent co-occurrences to identify global high-order label correlations. Second, we perform high-order disambiguation by selecting semantically relevant correlations and propagating their structural evidence to constituent labels for confidence refinement. Specifically, for each instance, we compute semantic alignment between the instance and each identified correlation, employ differentiable Gumbel-Softmax [Jang *et al.*, 2017] pruning to select relevant correlations, and broadcast their accumulated evidence to individual labels. Labels supported by multiple relevant correlations receive amplified confidence for disambiguation.

In summary, the main contributions are as follows:

- We propose to exploit high-order label correlations for PML disambiguation by mining frequent co-occurrence patterns.
- We design a differentiable pruning mechanism to select instance-relevant correlations and broadcast their structural evidence to constituent labels for confidence refinement.
- Extensive experiments on both real-world PML and UCI multi-label datasets demonstrate the effectiveness of the proposed method.

2 Related Work

In this section, we briefly review the most relevant works on PML. Existing methods can be broadly categorized into two main paradigms according to their optimization strategies: two-stage methods and unified methods.

Two-stage methods decompose the PML problem into sequential stages of label disambiguation followed by classifier training [Xie and Huang, 2018; Fang and Zhang, 2019; Zhang and Fang, 2021]. Confidence-based disambiguation strategies estimate the relevance of each candidate label before model training. For example, PARTICLE [Zhang and Fang, 2021] develops a probabilistic model that iteratively updates label confidences through expectation-maximization. PML-LRS [Sun *et al.*, 2019] utilizes low-rank and sparse decomposition to recover ground-truth label matrices from noisy candidate sets. Other feature-induced methods [Yu *et al.*, 2018; Xu *et al.*, 2024] leverage instance features to guide disambiguation. Meta-learning strategies have also been introduced, as exemplified by PML-MD [Xie *et al.*, 2021] which adaptively estimates confidences using meta-learning techniques. Other methods comprise multi-subspace representation [Li *et al.*, 2020], self-representation models [Lyu *et al.*, 2023], label distribution recovery [Xu *et al.*, 2020; Xu *et al.*, 2021; Lyu *et al.*, 2020; Kou *et al.*, 2025], and noisy label identification [Xie and Huang, 2021; Lyu *et al.*, 2021].

Unified methods attempt joint optimization to enable mutual reinforcement between confidence estimation and classifier training [Wang *et al.*, 2019; Li and Wang, 2020; Yan *et al.*, 2021]. Progressive enhancement methods iteratively refine pseudo-label distributions through alternating optimization [Xu *et al.*, 2021]. Alternative paradigms include adversarial training [Yan and Guo, 2021], mutual information maximization [Gong *et al.*, 2021], and self-distillation mechanisms [Xu *et al.*, 2022; Gong *et al.*, 2022]. Recently, methods exploiting label correlations have gained prominence [Zhao *et al.*, 2022; Li *et al.*, 2022; Kou *et al.*, 2024]. For example, WPML^{3}CP [Li *et al.*, 2024] captures label correlations from both measuring and modeling perspectives by employing Wasserstein distance. GLC [Sun *et al.*, 2022] integrates global and local label correlation through low-rank representation and label manifold regularization. Recently, graph-based methods [Chen *et al.*, 2020; Wang *et al.*, 2023; Hang and Zhang, 2023; He *et al.*, 2019] have also emerged as effective paradigms, and multi-view methods [Chen *et al.*, 2020; Wu *et al.*, 2020; Liu *et al.*, 2024; Wen *et al.*, 2024; Sun *et al.*, 2021] have been developed to handle scenarios with multiple feature representations.

Despite considerable progress, existing PML methods primarily focus on pairwise label correlations for disambiguation. However, we observe that high-order label correlations can provide even stronger disambiguation evidence in partial settings. Motivated by this, we propose to model high-order label correlations and design a broadcasting mechanism that propagates structural evidence from high-order patterns to individual labels for effective disambiguation.

3 The Proposed Method

In this section, we introduce the proposed REHC-PML method.

Task definition. Formally, given a training set $\mathcal{D} = (\mathbf{x}_i, S_i)_{i=1}^n$ with features $\mathbf{x}_i \in \mathbb{R}^d$ and candidate label sets $S_i \subseteq \{1, \ldots, l\}$, the unobserved ground-truth labels $\mathcal{Y}_i^* \subseteq \{1, \ldots, l\}$ satisfy $\mathcal{Y}_i^* \subseteq S_i$ for all i. The objective of partial multi-label learning is to learn a predictor $f : \mathbb{R}^d \to [0, 1]^l$ from $\mathcal{D}$ that correctly recovers the labels for unseen instances.

3.1 Overview of REHC-PML

As established in Section 1, high-order label correlations offer superior disambiguation evidence: ground-truth correlations manifest as frequent co-occurrences across candidate sets, while noise-induced patterns remain sparse. To exploit this property, we propose REHC-PML. As illustrated in Algorithm 1, the framework operates in two phases. In the first phase, we mine frequent high-order co-occurrences from candidate sets to identify global high-order label correlations. In the second phase, we adopt a self-training strategy that iteratively performs disambiguation and trains the classifier.

3.2 Global High-Order Correlations Construction

In the first phase, we construct a global set $\mathcal{K}$ of high-order label correlations by mining frequent co-occurrence patterns from candidate sets. The intuition is that when multiple ground-truth labels share correlations, they co-occur frequently across candidate sets, while random noise generates sporadic combinations. By identifying frequently co-occurring

labels, we can capture potential high-order label correlations for disambiguation.

We begin by enumerating all possible label co-occurrences up to a maximum order $p_{\max}$. Given the candidate label set S_i for instance $\mathbf{x}_i$, we enumerate all label subsets

$$\mathcal{K}_i = \{\sigma \subseteq S_i : 1 \leq |\sigma| \leq p_{\max}\}, \tag{1}$$

where σ denotes a potential high-order label correlation and $|\sigma|$ denotes the number of labels involved (the *order* of the correlation [Huang *et al.*, 2017; Zhang and Zhang, 2010]).

Then, we aggregate these correlations across all instances and compute the global frequency for each unique correlation

$$\text{freq}(\sigma) = \sum_{i=1}^{n} \mathbb{I}(\sigma \subseteq S_i), \tag{2}$$

where $\mathbb{I}(\cdot)$ is the indicator function. This frequency measures how often the label subset σ appears across candidate sets. High frequency indicates that the labels in σ consistently co-occur, suggesting they may share a ground-truth high-order correlation.

To focus on recurrent co-occurrence patterns and reduce computational complexity, we retain only the most frequent label subsets for each order. For each order $p \in \{1, \ldots, p_{\max}\}$, we select the top-$K_p$ label subsets with the highest frequencies. This filtering step prunes extremely rare correlations while preserving more frequent ones. The result is a global set of high-order label correlations

$$\mathcal{K} = \{\sigma_1, \ldots, \sigma_M\}, \tag{3}$$

where $M = \sum_{p=1}^{p_{\max}} K_p$ denotes the total number of retained correlations. Each $\sigma_j \in \mathcal{K}$ represents a high-order label correlation identified through frequent co-occurrence, serving as potential structural evidence for disambiguation.

3.3 Self-Training with High-Order Propagation

In the second phase, we adopt a self-training strategy to optimize the framework, which iteratively generates pseudo-labels and trains the classifier on these refined targets. The key challenge lies in how to generate high-quality pseudo-labels that effectively distinguish ground-truth labels from noise. To this end, we design a high-order guided propagation mechanism that leverages the structural evidence from the global set of high-order label correlations $\mathcal{K}$ to refine label confidence.

High-Order Guided Pseudo-Label Generation

Although the global set $\mathcal{K}$ captures high-order label correlations through frequent co-occurrence patterns, not all correlations in $\mathcal{K}$ are relevant to every instance. We design a high-order guided propagation mechanism that dynamically selects instance-relevant correlations and broadcasts their structural evidence to boost the confidence of constituent labels.

Semantic Relevance Computation. To determine which correlations $\sigma_j \in \mathcal{K}$ are relevant to a instance $\mathbf{x}_i$, we compute a relevance score measuring their semantic alignment as

$$s_{i,j} = \frac{g(\mathbf{x}_i)^\top \mathbf{h}_{\sigma_j}}{\sqrt{d_e}}, \tag{4}$$

Algorithm 1 Training Procedure of REHC-PML

Input: Training set $\mathcal{D} = \{(\mathbf{x}_i, S_i)\}_{i=1}^n$, max order $p_{\max}$, top-K thresholds $\{K_p\}_{p=1}^{p_{\max}}$, self-training probability ρ, confidence threshold θ, number of epochs T
Output: Trained classifier f

// Phase 1: Construct Global High-Order Correlations
for $i = 1$ to n **do**
  Enumerate co-occurrences $\mathcal{K}_i$ via Eq.(1)
end for
Compute frequencies freq(σ) for all unique σ via Eq.(2)
for $p = 1$ to $p_{\max}$ **do**
  Retain top-K_p frequent order-p correlations
end for
Construct global set $\mathcal{K} = \{\sigma_1, \ldots, \sigma_M\}$
Build membership matrix $\mathbf{C} \in \{0, 1\}^{M \times l}$

// Phase 2: Self-Training with High-Order Propagation
Initialize classifier f, projection network g, label embeddings $\mathbf{E}$, auxiliary network h
for $t = 1$ to T **do**
  for each mini-batch $\mathcal{B} \subset \mathcal{D}$ **do**
    Compute correlation representations $\{\mathbf{h}_{\sigma_j}\}_{j=1}^M$ via Eq.(5)
    for each $(\mathbf{x}_i, S_i) \in \mathcal{B}$ **do**
      Compute relevance scores $\{s_{i,j}\}_{j=1}^M$ via Eq.(4)
      Generate pruning mask $\mathbf{m}_i$ via Eq.(6)
      Compute structural support $\mathbf{v}_i = \mathbf{m}_i \mathbf{C}$
      Refine label confidence $\{\tilde{y}_{i,k}\}_{k=1}^l$ via Eq.(8)
    end for
    if rand() $< \rho$ **then**
      Generate pseudo-labels $\hat{Y}_{i,k}$ via Eq.(9)
      Compute loss $\mathcal{L}$ via Eq.(10) with pseudo-labels
    else
      Compute loss $\mathcal{L}$ via Eq.(10) with candidate labels only
    end if
    Update all parameters via gradient descent on $\mathcal{L}$
  end for
end for
return Trained classifier f

where $g(\mathbf{x}_i) \in \mathbb{R}^{d_e}$ is the instance representation obtained via a non-linear projection network $g : \mathbb{R}^d \to \mathbb{R}^{d_e}$ (implemented as a multi-layer perceptron) [Bengio *et al.*, 2013], d_e denotes the embedding dimension, and $\mathbf{h}_{\sigma_j} \in \mathbb{R}^{d_e}$ is the correlation representation. The representation of σ_j is computed as a weighted sum of its constituent label embeddings as

$$\mathbf{h}_{\sigma_j} = \sum_{k \in \sigma_j} \frac{\exp(f_{\text{attn}}(\mathbf{E}_k))}{\sum_{k' \in \sigma_j} \exp(f_{\text{attn}}(\mathbf{E}_{k'}))} \mathbf{E}_k, \tag{5}$$

where $\mathbf{E} \in \mathbb{R}^{l \times d_e}$ is a learnable label embedding matrix and $f_{\text{attn}} : \mathbb{R}^{d_e} \to \mathbb{R}$ is a learned attention function [Vaswani *et al.*, 2017] that identifies the most representative labels within each correlation. A high relevance score indicates that the high-order correlation represented by σ_j strongly correlates with the semantic content of the instance.

Differentiable Correlation Pruning. Based on the relevance scores, we prune irrelevant correlations from $\mathcal{K}$ to derive an instance-specific subset $\mathcal{K}_i \subseteq \mathcal{K}$. To implement this pruning in a differentiable manner, we employ the Gumbel-Softmax relaxation [Jang *et al.*, 2017; Maddison *et al.*, 2017]

as a learnable gating mechanism. The pruning mask $\mathbf{m}_i \in \{0,1\}^M$ is generated as

$$\mathbf{m}_{i,j} = \text{STE}\left(\varsigma\left(\frac{s_{i,j}+\epsilon_j}{\tau}\right)\right), \tag{6}$$

where $\varsigma(\cdot)$ denotes the sigmoid function, $\epsilon_j = -\log(-\log(u_j))$ with $u_j \sim \text{Uniform}(0,1)$ denotes the Gumbel noise, $\tau > 0$ is the temperature parameter, and $\text{STE}(\cdot)$ denotes the Straight-Through Estimator that enables gradient flow through the discrete sampling. Here $\mathbf{m}_{i,j} = 1$ indicates that correlation σ_j is selected for instance i, while $\mathbf{m}_{i,j} = 0$ signifies that the correlation is pruned. This effectively reduces $\mathcal{K}$ to a sparse, instance-adaptive subset of relevant correlations.

Correlation-to-Label Broadcasting. With the pruning mask $\mathbf{m}_i$, we propagate the retained high-order correlations to individual labels for confidence refinement. Intuitively, if a label is contained in multiple retained correlations, it receives stronger support and is more likely to be a ground-truth label. To quantify this support, we compute a structural vote vector $\mathbf{v}_i \in \mathbb{R}^l$, where $\mathbf{v}_{i,k}$ counts how many retained correlations contain label k, representing the accumulated structural evidence from high-order label correlations as

$$\mathbf{v}_i = \mathbf{m}_i \mathbf{C}, \tag{7}$$

where $\mathbf{C} \in \{0,1\}^{M \times l}$ is a composition matrix encoding membership relationships, with $\mathbf{C}_{j,k} = 1$ if label $k \in \sigma_j$ and 0 otherwise.

Then, we use this structural signal to refine the label confidence. Specifically, we combine $\mathbf{v}_i$ with a base confidence $\hat{y}^{\text{base}}_{i,k}$ estimated by an auxiliary network from instance features as follows

$$\tilde{y}_{i,k} = \hat{y}^{\text{base}}_{i,k} \cdot \big(1 + \beta \cdot \varsigma(\mathbf{v}_{i,k} - \gamma)\big), \tag{8}$$

where γ controls the activation threshold, and β controls the boosting intensity. This formulation ensures that only labels supported by sufficient retained combinations (*i.e.*, $\mathbf{v}_{i,k} > \gamma$) receive meaningful amplification, effectively distinguishing ground-truth labels from noise.

Classifier Training with Pseudo-Labels
The refined confidence scores are converted to hard pseudo-labels according to

$$\hat{Y}_{i,k} = \mathbb{I}(\tilde{y}_{i,k} > \theta) \cdot \mathbb{I}(k \in S_i), \tag{9}$$

where θ is the confidence threshold and $\mathbb{I}(k \in S_i)$ ensures that the pseudo-labels remain within the candidate set. The classifier is then trained to align with these refined targets.

The classification loss $\mathcal{L}$ combines consistency with candidate labels and alignment with pseudo-labels as

$$\mathcal{L} = \frac{1}{nl}\sum_{i=1}^{n}\sum_{k=1}^{l}\Big[\ell_{\text{BCE}}\big(f(\mathbf{x}_i)_k, \mathbb{I}(k \in S_i)\big) + \lambda \cdot \ell_{\text{BCE}}\big(f(\mathbf{x}_i)_k, \hat{Y}_{i,k}\big)\Big], \tag{10}$$

where $f(\mathbf{x}_i)_k$ denotes the classifier output for label k, and λ balances the two terms. To balance optimization stability and computational efficiency, the self-training phase (i.e., including the pseudo-label term) is stochastically activated with probability ρ at each iteration. The complete training procedure is summarized in Algorithm 1.

Dataset	n	d	l	**#AvgLabels**
Real-world PML Datasets				
mirflickr	10,433	100	7	1.77
music_emotion	6,833	98	11	2.42
music_style	6,839	98	10	1.44
YeastBP	6,139	6,139	217	5.54
YeastCC	6,139	6,139	50	1.35
YeastMF	6,139	6,139	39	1.01
UCI Multi-label Datasets				
bibtex	7,395	1,836	159	2.40
bookmarks	87,856	2,150	208	2.03
bow-train	22,000	100	101	4.29
computers	12,444	34,096	33	1.51
delicious	16,105	500	983	19.02
enron	1,702	1,001	53	3.38
entertainment	2,545	640	21	1.42
nuswide	133,441	500	81	1.87

Table 1: Statistics of benchmark datasets used in experiments. n: number of instances; d: feature dimensionality; l: number of labels; #AvgLabels: average number of labels per instance.

4 Experiments

In this section, we conduct comprehensive experiments to evaluate the effectiveness of the proposed REHC-PML method.

4.1 Experimental Settings

Datasets. We evaluate REHC-PML on six real-world PML datasets and eight UCI multi-label datasets. The real-world datasets include *mirflickr*, *music_emotion*, *music_style*, *YeastBP*, *YeastCC*, and *YeastMF*, which inherently contain candidate label annotations from practical scenarios such as image annotation and biological function prediction. For the UCI multi-label datasets, we include *bibtex*, *bookmarks*, *bow-train*, *computers*, *delicious*, *enron*, *entertainment*, and *nuswide*. Detailed characteristics of all datasets are summarized in Table 1.

To simulate the partial multi-label learning scenario on UCI datasets, we construct candidate label sets by augmenting the ground-truth labels with randomly sampled irrelevant labels following the standard protocol [Zhang and Fang, 2021; Xie *et al.*, 2021]. Specifically, for each instance $\mathbf{x}_i$, we add $m\% \cdot \|\mathbf{y}_i^*\|_0$ irrelevant labels to the ground-truth label set $\mathbf{y}_i^*$ to form the candidate label set S_i, where $m \in \{50, 100, 150\}$ controls the noise ratio. All experiments are conducted with 5-fold cross-validation, and we report the mean and standard deviation across folds.

Baseline Methods. We compare REHC-PML against nine state-of-the-art PML methods covering both two-stage and unified paradigms: PARTICLE [Zhang and Fang, 2021], PML-NI [Xie and Huang, 2021], PML-MD [Xie *et al.*, 2021], GLC [Sun *et al.*, 2022], PAMB [Liu *et al.*, 2023], PLAIN [Wang *et al.*, 2023], PARD [Hang and Zhang, 2023], WPML^{3}CP [Li *et al.*, 2024], and PML-FSMIR [Gao *et al.*, 2025]. For fair comparison, all baseline methods are implemented using their publicly available code with recommended hyperparameter settings. Detailed descriptions of baseline methods are provided in the Appendix A[1].

Datasets	REHC-PML	PML-FSMIR	WPML^{3}CP	PLAIN	PARD	PAMB	GLC	PMLMD	PMLNI	PARTICLE
					Average Precision ↑					
mirflickr	**.868±.006**	.770±.011	.856±.006	.846±.009	.704±.048	.722±.074	.750±.007	.842±.017	.791±.008	.781±.176
music_emotion	.657±.005	.614±.005	.623±.004	**.658±.008**	.559±.006	.640±.005	.505±.007	.633±.005	.608±.008	.507±.007
music_style	**.754±.008**	.699±.013	.703±.009	.735±.006	.693±.006	.715±.007	.652±.008	.714±.015	.736±.008	.654±.003
YeastBP	**.618±.008**	.505±.011	.552±.009	.433±.013	.506±.010	.353±.205	.467±.017	.589±.008	.581±.015	.382±.197
YeastCC	**.860±.004**	.791±.005	.836±.007	.824±.008	.833±.004	.653±.134	.791±.016	.842±.004	.830±.006	.678±.117
YeastMF	**.794±.005**	.726±.007	.779±.009	.760±.012	.748±.006	.636±.111	.703±.005	.785±.003	.765±.009	.671±.102
					Coverage Error ↓					
mirflickr	**.196±.005**	.281±.007	.209±.004	.203±.005	.284±.014	.274±.047	.300±.005	.201±.011	.228±.004	.253±.059
music_emotion	.390±.004	.432±.006	.405±.001	**.387±.005**	.444±.007	.407±.005	.506±.005	.420±.008	.409±.006	.505±.007
music_style	**.192±.005**	.233±.009	.215±.007	.209±.003	.219±.004	.214±.004	.295±.006	.213±.010	.200±.005	.292±.008
YeastBP	**.223±.011**	.340±.010	.281±.005	.420±.006	.311±.014	.561±.193	.326±.017	.231±.005	.293±.017	.436±.153
YeastCC	**.073±.008**	.131±.009	.09±.007	.088±.004	.092±.004	.272±.122	.125±.007	.081±.005	.104±.003	.222±.098
YeastMF	**.092±.006**	.141±.001	.104±.009	.099±.006	.122±.004	.251±.094	.149±.002	.096±.003	.123±.008	.204±.080
					Hamming Loss ↓					
mirflickr	**.149±.003**	.180±.004	.158±.004	.164±.005	.252±.025	.224±.034	.181±.003	.177±.009	.196±.004	.177±.039
music_emotion	.211±.003	.223±.002	**.208±.003**	.210±.003	.336±.005	.220±.002	.256±.004	.217±.004	.227±.004	.256±.003
music_style	**.108±.002**	.119±.003	.125±.001	.111±.002	.829±.006	.119±.003	.125±.002	.120±.005	.113±.003	.125±.002
YeastBP	.036±.001	.041±.001	.031±.001	.043±.001	**.025±.002**	.048±.021	.042±.001	.037±.001	.037±.001	.047±.021
YeastCC	.037±.001	.040±.001	.023±.001	.039±.001	**.026±.001**	.042±.016	.040±.001	.038±.001	.038±.001	.041±.015
YeastMF	.039±.006	.046±.001	.030±.002	.037±.010	**.026±.001**	.040±.020	.039±.012	.039±.006	.033±.011	.039±.019
					One Error ↓					
mirflickr	**.134±.008**	.244±.015	.145±.009	.187±.018	.475±.136	.417±.101	.248±.010	.232±.031	.300±.014	.255±.265
music_emotion	**.380±.014**	.440±.013	.457±.009	.388±.017	.563±.007	.407±.010	.593±.010	.425±.017	.479±.021	.593±.014
music_style	**.318±.010**	.375±.017	.392±.049	.336±.009	.405±.009	.373±.016	.405±.009	.381±.024	.346±.012	.405±.005
YeastBP	**.663±.016**	.778±.016	.717±.016	.851±.009	.771±.011	.974±.020	.847±.015	.696±.019	.696±.010	.947±.028
YeastCC	**.757±.019**	.825±.014	.777±.003	.798±.015	.781±.016	.974±.006	.815±.008	.775±.014	.783±.007	.957±.028
YeastMF	**.820±.008**	.893±.007	.838±.006	.860±.006	.869±.008	.974±.012	.911±.006	.829±.007	.848±.009	.959±.035
					Ranking Loss ↓					
mirflickr	**.079±.004**	.180±.009	.094±.004	.089±.005	.195±.019	.179±.075	.197±.006	.093±.013	.121±.005	.147±.128
music_emotion	**.221±.004**	.268±.005	.242±.003	.222±.005	.287±.006	.234±.002	.362±.006	.244±.007	.246±.005	.361±.009
music_style	**.132±.005**	.172±.012	.155±.008	.149±.001	.161±.005	.153±.004	.230±.006	.153±.010	.139±.005	.231±.006
YeastBP	.121±.004	.180±.006	.146±.002	.217±.006	.169±.006	.391±.121	.198±.011	**.112±.003**	.155±.011	.297±.099
YeastCC	**.051±.005**	.094±.003	.063±.004	.065±.003	.065±.002	.230±.105	.089±.006	.056±.003	.074±.003	.174±.078
YeastMF	**.080±.005**	.118±.002	.088±.008	.087±.005	.105±.005	.237±.081	.132±.002	**.080±.004**	.107±.005	.186±.065

Table 2: Experimental results (mean±std) on real-world PML datasets. The best performance is shown in **boldface**.

Evaluation Metrics. Following standard practice in multi-label learning [Zhang and Zhou, 2014; Liu *et al.*, 2021], we adopt five widely-used evaluation metrics: *Average Precision (AP)*, *Coverage Error (CE)*, *Hamming Loss (HL)*, *One Error (OE)*, and *Ranking Loss (RL)* [Fürnkranz *et al.*, 2008]. The symbols ↑ and ↓ indicate that larger and smaller values correspond to better performance, respectively.

4.2 Results on Real-world Datasets

Table 2 presents the experimental results on six real-world PML datasets. The best performance is highlighted in boldface. We observe that REHC-PML consistently achieves superior or competitive performance across all datasets and evaluation metrics. Specifically, REHC-PML achieves the best Average Precision on five out of six datasets, with notable improvements on *mirflickr*, *YeastBP*, and *YeastMF* compared to other baseline methods. The only exception is the *music_emotion* dataset, where PLAIN performs slightly better. These results indicate that modeling high-order label correlations through frequent co-occurrence mining effectively captures the inherent structural patterns in candidate label sets, leading to more accurate label disambiguation. In terms of ranking-based metrics (*i.e.*, Coverage Error, One Error, and Ranking Loss), REHC-PML consistently outperforms baseline methods on most datasets, especially on those with complex label structures such as *YeastBP* and *YeastCC*, demonstrating the success of the high-order to low-order broadcasting mechanism in label ranking. Regarding Hamming Loss, WPML^{3}CP achieves competitive results, particularly on the Yeast datasets, due to its Wasserstein distance-based formulation optimized for dense label distributions. However, REHC-PML remains competitive and achieves the best Hamming Loss on *mirflickr* and *music_style*. Additionally, some baseline methods (*i.e.*, PARTICLE and PAMB) exhibit high variance across different datasets, while REHC-PML demonstrates stable performance with consistently low variance, suggesting that the proposed high-order correlation-based framework provides robust structural guidance for label disambiguation.

4.3 Results on UCI Datasets

To further evaluate the robustness of REHC-PML under varying noise levels, we conduct experiments on UCI multi-label datasets with noise ratios $m \in \{50, 100, 150\}$. Due to space constraints, we present representative results for selected datasets with $m = 150$ in Table 3. Complete results are provided in Appendix B[1]. The selected UCI datasets are particularly challenging due to their large size or high label dimensionality. These characteristics pose significant challenges for PML methods in terms of both disambiguation accuracy and computational time. Notably, entries marked with "-" indicate that the corresponding methods failed to produce results

[1]https://github.com/yuanchaodai/IJCAI2026_REHC-PML_Appendix

Datasets	REHC-PML	PML-FSMIR	WPML^{3}CP	PLAIN	PARD	PAMB	GLC	PMLMD	PMLNI	PARTICLE
					Average Precision ↑					
bibtex	**.513±.009**	.500±.009	.455±.052	.298±.009	.504±.016	.149±.013	.285±.006	.455±.008	.510±.011	.271±.006
bookmarks	**.489±.002**	-	.420±.002	.272±.003	.421±.003	-	.423±.002	.345±.015	.465±.002	-
bow-train	**.757±.005**	.669±.004	.679±.003	.700±.006	.678±.002	.488±.023	.651±.002	.674±.006	.712±.005	.693±.015
computers	**.748±.004**	-	.696±.007	.698±.003	.674±.009	-	.657±.006	.706±.006	.643±.006	-
delicious	**.399±.003**	.264±.004	.260±.003	.224±.005	.295±.002	-	.048±.004	-	.357±.001	-
enron	**.689±.015**	.627±.013	.613±.005	.660±.019	.681±.010	.471±.076	.576±.008	.684±.015	.495±.003	.597±.047
entertainment	**.771±.004**	-	.691±.034	.705±.004	.692±.005	-	.607±.007	.706±.007	.618±.009	-
nuswide	**.671±.001**	-	.602±.002	.636±.001	.584±.002	-	.569±.002	.586±.011	.637±.001	-
					Coverage Error ↓					
bibtex	.181±.005	.193±.007	.222±.045	.355±.020	**.157±.009**	.524±.005	.254±.007	.211±.006	.255±.005	.483±.005
bookmarks	**.156±.001**	-	.169±.004	.311±.003	.182±.009	-	.159±.001	.235±.004	.199±.001	-
bow-train	.185±.005	.198±.002	**.174±.005**	.283±.009	.202±.004	.573±.011	.207±.002	.249±.018	.203±.005	.200±.015
computers	.138±.007	-	.131±.007	.164±.004	**.116±.004**	-	.122±.003	.128±.004	.186±.004	-
delicious	**.549±.007**	.643±.005	.625±.003	.785±.011	.818±.021	-	.962±.004	-	.680±.006	-
enron	.277±.007	**.246±.007**	.324±.018	.290±.009	.262±.013	.506±.099	.277±.009	.252±.008	.433±.016	.311±.041
entertainment	.139±.015	-	.138±.014	.167±.005	**.132±.004**	-	.144±.005	.138±.007	.200±.006	-
nuswide	**.121±.001**	-	.142±.001	.179±.002	.166±.002	-	.159±.001	.181±.005	.152±.001	-
					Hamming Loss ↓					
bibtex	.017±.000	.017±.000	.017±.001	.022±.000	**.014±.000**	.025±.001	.023±.000	.018±.000	.017±.000	.024±.000
bookmarks	.013±.000	-	.019±.000	.016±.000	**.009±.000**	-	.014±.000	.015±.000	.013±.000	-
bow-train	**.034±.000**	.040±.000	.037±.000	.037±.000	.035±.000	.052±.003	.041±.000	.039±.000	.037±.000	.039±.002
computers	**.033±.001**	-	.041±.001	.036±.001	.038±.000	-	.040±.001	.036±.001	.040±.001	-
delicious	**.024±.000**	.028±.000	.028±.000	.029±.000	.024±.000	-	.037±.001	-	**.024±.000**	-
enron	.052±.002	.059±.001	.060±.000	.055±.002	**.048±.001**	.068±.008	.063±.001	.054±.001	.075±.001	.063±.008
entertainment	**.046±.001**	-	.060±.005	.054±.002	.058±.000	-	.073±.002	.055±.001	.065±.002	-
nuswide	.025±.000	-	.027±.000	.025±.000	**.023±.000**	-	.027±.000	.027±.001	.025±.000	-
					One Error ↓					
bibtex	.438±.018	.433±.011	.518±.056	.647±.013	.450±.019	.923±.012	.729±.010	.509±.013	**.399±.012**	.749±.019
bookmarks	**.533±.003**	-	.560±.002	.760±.004	.607±.003	-	.622±.003	.685±.020	.550±.001	-
bow-train	**.144±.006**	.202±.009	.175±.003	.171±.005	.205±.004	.237±.029	.219±.005	.215±.014	.180±.005	.178±.018
computers	**.286±.004**	-	.357±.008	.336±.005	.414±.015	-	.410±.009	.338±.010	.400±.010	-
delicious	**.303±.002**	.445±.011	.457±.016	.513±.013	.399±.011	-	.933±.006	-	.345±.006	-
enron	**.225±.027**	.314±.025	.280±.012	.254±.023	.227±.012	.640±.159	.300±.018	.242±.027	.448±.018	.308±.028
entertainment	**.273±.006**	-	.403±.057	.357±.008	.415±.005	-	.564±.010	.367±.007	.473±.011	-
nuswide	**.562±.003**	-	.631±.003	.585±.001	.652±.002	-	.672±.003	.641±.023	.595±.000	-
					Ranking Loss ↓					
bibtex	**.106±.004**	.110±.005	.135±.041	.233±.013	.108±.006	.369±.007	.163±.005	.117±.004	.142±.005	.327±.002
bookmarks	**.100±.002**	-	.143±.004	.223±.002	.117±.005	-	.101±.001	.157±.003	.124±.001	-
bow-train	**.052±.002**	.061±.001	**.052±.002**	.092±.004	.061±.001	.245±.010	.064±.001	.075±.004	.058±.002	.060±.006
computers	.088±.005	-	.086±.005	.112±.002	**.075±.003**	-	.079±.003	.083±.002	.132±.002	-
delicious	**.113±.002**	.149±.002	.140±.002	.195±.004	.182±.007	-	.670±.006	-	.141±.001	-
enron	**.098±.003**	**.098±.003**	.127±.011	.103±.005	.094±.004	.224±.043	.109±.006	.088±.003	.201±.005	.116±.016
entertainment	**.092±.013**	-	.100±.013	.120±.003	**.092±.003**	-	.108±.003	.098±.006	.154±.005	-
nuswide	**.062±.001**	-	.076±.000	.095±.002	.090±.001	-	.087±.001	.099±.004	.081±.001	-

Table 3: Experimental results (mean±std) on eight UCI PML datasets ($m = 150$). The best performance is shown in **boldface**.

Method	AP	CE	HL	OE	RL	Total
PML-FSMIR	18/0/0	15/0/3	15/3/0	17/0/1	16/1/1	81/4/5
WPML^{3}CP	30/0/0	24/0/6	24/2/4	30/0/0	27/2/1	135/4/11
PLAIN	29/0/1	29/0/1	25/3/2	30/0/0	30/0/0	143/3/4
PARD	30/0/0	20/0/10	11/1/18	29/0/1	21/2/7	111/3/36
PAMB	15/0/0	15/0/0	15/0/0	15/0/0	15/0/0	75/0/0
GLC	30/0/0	27/1/2	29/1/0	30/0/0	29/0/1	145/2/3
PML-MD	26/1/0	18/0/9	23/4/0	26/0/1	17/3/7	110/8/17
PML-NI	28/1/1	30/0/0	17/10/3	27/0/3	30/0/0	132/11/7
PARTICLE	15/0/0	15/0/0	14/1/0	15/0/0	15/0/0	74/1/0

Table 4: Win/tie/loss counts of pairwise t-test between REHC-PML and each comparing method at 0.05 significance level.

within reasonable time constraints due to their computational complexity on these large-scale datasets.

From Table 3, we observe that REHC-PML achieves consistently superior performance across all eight datasets. Most notably, REHC-PML ranks first in Average Precision on all datasets, with substantial improvements on *delicious* (+11.8% over the second-best PMLNI) and *bookmarks* (+5.2% over PMLNI). For the Coverage Error and Ranking Loss, PARD achieves slightly better results on *computers*, while REHC-PML excels on most datasets. We attribute this to the relatively small average label cardinality of these datasets, where pairwise label correlations may suffice for effective ranking. However, on datasets with richer label structures such as *delicious* (19.02 average labels), REHC-PML significantly outperforms PARD, confirming the advantage of high-order label correlations for complex PML scenarios.

4.4 Statistical Analysis

To further substantiate the effectiveness of REHC-PML, we conduct pairwise t-tests between our proposed method and each baseline across 30 datasets and 5 evaluation metrics (150 pairwise comparisons per baseline). It is worth noting that incomplete results (denoted as '-' in the experimental tables) are excluded from the statistical comparisons. Table 4 summarizes the win/tie/loss counts at the 0.05 significance level. The results demonstrate that REHC-PML achieves statistically significant improvements in the overwhelming majority of

Variant	(a)	(b)	(c)	mirflickr	music_emotion	music_style	YeastBP	YeastCC	YeastMF
I		✓	✓	0.846±0.012	0.634±0.008	0.724±0.009	0.608±0.013	0.851±0.007	0.788±0.004
II	✓		✓	0.856±0.007	0.630±0.020	0.752±0.007	0.608±0.006	0.859±0.006	0.793±0.005
III	✓	✓		0.853±0.009	0.638±0.015	0.740±0.012	0.607±0.007	0.855±0.005	0.791±0.007
IV	✓	✓	✓	**0.868±0.006**	**0.657±0.005**	**0.754±0.008**	**0.618±0.008**	**0.860±0.004**	**0.794±0.005**

Table 5: Ablation study results (Average Precision) on real-world PML datasets.

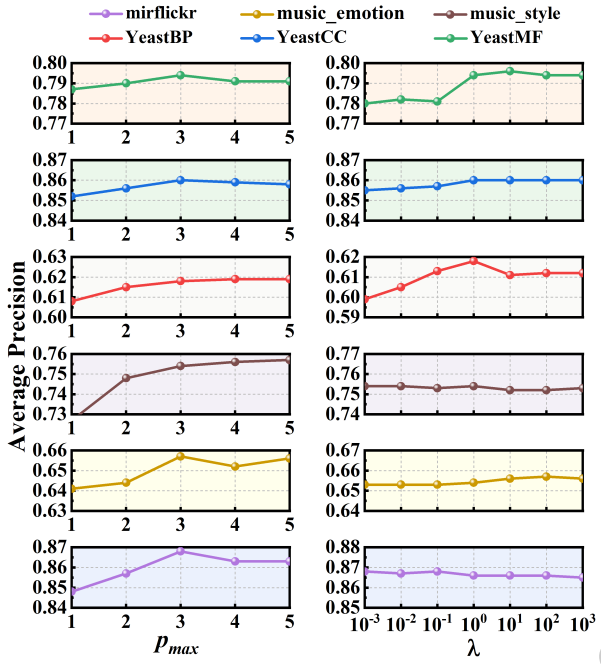

Figure 2: Parameter sensitivity analysis of parameters $\{p_{max}, \lambda\}$ on six real-world PML datasets.

Evaluation metric	F_F	Critical value
Average Precision	20.920	
Coverage Error	11.171	
Hamming Loss	13.680	1.955
One Error	15.025	
Ranking Loss	12.629	

Table 6: Friedman statistics F_F in terms of each evaluation metric and the critical value at 0.05 significance level (#comparing algorithms $k = 10$, #datasets $N = 15$).

comparisons. Specifically, REHC-PML significantly outperforms GLC in 96.7% (145/150) of tests, PLAIN in 95.3% (143/150) of tests, and PML-NI in 88.0% (132/150) of tests. Against the state-of-the-art method PML-FSMIR, REHC-PML wins in 90.0% (81/90) of comparisons.

Furthermore, the Friedman test is used to statistically compare the relative performance among all competing methods. Table 6 reports the Friedman statistic F_F in terms of each evaluation metric. The results demonstrate that all five Friedman statistics substantially exceed the critical value of 1.955 at the 0.05 significance level. Average Precision exhibits the highest discriminative power ($F_F = 20.920$), followed by One Error ($F_F = 15.025$), Hamming Loss ($F_F = 13.680$), Ranking Loss ($F_F = 12.629$), and Coverage Error ($F_F = 11.171$). Notably, REHC-PML consistently achieves the best mean rank across all metrics, demonstrating that our proposed high-order correlation-based framework delivers robust and reliable performance improvements over existing PML baselines.

4.5 Ablation Study

To validate the contribution of different components, we conduct an ablation study on six real-world PML datasets. Table 5 presents the Average Precision for four variants, where we evaluate three key components: (a) high-order label correlation structure, (b) attention-based semantic aggregation, and (c) Gumbel-Softmax structural pruning. The results show that each component contributes to the overall performance. Specifically, Variant I leads to the most significant performance degradation, indicating that modeling high-order label correlations is crucial for effective label disambiguation. The Gumbel-Softmax structural pruning also contributes positively (Variant III), particularly on *music_emotion* and *music_style*, demonstrating that differentiable correlation selection effectively filters out noisy label co-occurrence patterns. These results collectively validate the effectiveness of each proposed component in the REHC-PML framework.

4.6 Parameter Analysis

We investigate the sensitivity of two key hyperparameters: the maximum correlation order $p_{\max}$ and the pseudo-label loss weight λ. As shown in Figure 2, the left column demonstrates that performance generally improves as $p_{\max}$ increases from 1 to 3, confirming that modeling high-order label correlations enhances disambiguation. The right column shows that λ exhibits relatively stable performance across a wide range (10^{-2} to 10^{2}), with optimal values typically around $\lambda = 1$. This suggests that our method is robust to the balance between candidate label consistency and pseudo-label alignment.

5 Conclusion

This paper proposes REHC-PML to address the limitation of existing PML methods that neglect high-order label correlations. By mining high-confidence label combinations and broadcasting them to constituent labels, our method provides structural guidance for disambiguation. Experiments on six real-world and eight UCI datasets demonstrate superior performance across multiple metrics, particularly on datasets with complex label structures. The results validate that high-order co-occurrence patterns offer more effective disambiguation than instance-level or pairwise methods. Future work will explore more effective structural pruning methods and higher-order optimization techniques.

Acknowledgments

We would like to acknowledge support for this project from the National Natural Science Foundation of China (No.62276113) and the Graduate Innovation Fund of Jilin University (No.2026CX208).

References

[Bengio *et al.*, 2013] Yoshua Bengio, Aaron Courville, and Pascal Vincent. Representation learning: A review and new perspectives. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(8):1798–1828, 2013.

[Chen *et al.*, 2020] Zesen Chen, Xuan Wu, Qingguo Chen, Yao Hu, and Minling Zhang. Multi-view partial multi-label learning with graph-based disambiguation. In *Proceedings of the Association for the Advancement of Artificial Intelligence*, pages 3553–3560, 2020.

[Fang and Zhang, 2019] Jun-Peng Fang and Min-Ling Zhang. Partial multi-label learning via credible label elicitation. In *Proceedings of the Association for the Advancement of Artificial Intelligence*, pages 3518–3525, 2019.

[Fürnkranz *et al.*, 2008] Johannes Fürnkranz, Eyke Hüllermeier, Eneldo Loza Mencía, and Klaus Brinker. Multilabel classification via calibrated label ranking. *Machine Learning*, 73(2):133–153, 2008.

[Gao *et al.*, 2025] Wanfu Gao, Hanlin Pan, Qingqi Han, and Kunpeng Liu. Noise-resistant label reconstruction feature selection for partial multi-label learning. In *Proceedings of the International Joint Conference on Artificial Intelligence*, pages 5172–5180, 2025.

[Gong *et al.*, 2021] Xiuwen Gong, Dong Yuan, and Wei Bao. Understanding partial multi-label learning via mutual information. In *Proceedings of the Neural Information Processing Systems*, pages 4147–4156, 2021.

[Gong *et al.*, 2022] Xiuwen Gong, Dong Yuan, and Wei Bao. Partial multi-label learning via large margin nearest neighbour embeddings. In *Proceedings of the Association for the Advancement of Artificial Intelligence*, pages 6729–6736, 2022.

[Hang and Zhang, 2023] Jun-Yi Hang and Min-Ling Zhang. Partial multi-label learning with probabilistic graphical disambiguation. In *Proceedings of the Neural Information Processing Systems*, 2023.

[He *et al.*, 2019] Shuo He, Ke Deng, Li Li, Senlin Shu, and Li Liu. Discriminatively relabel for partial multi-label learning. In *Proceedings of the IEEE International Conference on Data Mining*, pages 280–288, 2019.

[Huang *et al.*, 2017] Jun Huang, Guorong Li, Shuhui Wang, Zhe Xue, and Qingming Huang. Multi-label classification by exploiting local positive and negative pairwise label correlation. *Neurocomputing*, 257:164–174, 2017.

[Jang *et al.*, 2017] Eric Jang, Shixiang Gu, and Ben Poole. Categorical reparameterization with gumbel-softmax. In *Proceedings of the International Conference on Learning Representations*, 2017.

[Kou *et al.*, 2024] Zhiqiang Kou, Jing Wang, Jiawei Tang, Yuheng Jia, Boyu Shi, and Xin Geng. Exploiting multi-label correlation in label distribution learning. In *International Joint Conference on Artificial Intelligence*, pages 4326–4334, 2024.

[Kou *et al.*, 2025] Zhiqiang Kou, Haoyuan Xuan, Jingyu Zhu, Hailin Wang, Ming-kun Xie, Changwei Wang, Jing Wang, Yuheng Jia, and Xin Geng. Tail-aware reconstruction of incomplete label distributions with low-rank and sparse modeling. *IEEE Transactions on Circuits and Systems for Video Technology*, pages 1–1, 2025.

[Li and Wang, 2020] Ximing Li and Yang Wang. Recovering accurate labeling information from partially valid data for effective multi-label learning. In *Proceedings of the International Joint Conference on Artificial Intelligence*, pages 1373–1380, 2020.

[Li *et al.*, 2020] Ziwei Li, Gengyu Lyu, and Songhe Feng. Partial multi-label learning via multi-subspace representation. In *Proceedings of the International Joint Conference on Artificial Intelligence*, pages 2612–2618, 2020.

[Li *et al.*, 2022] Feng Li, Shengfei Shi, and Hongzhi Wang. Partial multi-label learning via specific label disambiguation. *Knowledge-Based Systems*, 250:109093, 2022.

[Li *et al.*, 2024] Ximing Li, Yuanchao Dai, Bing Wang, Changchun Li, Renchu Guan, Fangming Gu, and Jihong Ouyang. WPML^{3}CP: wasserstein partial multi-label learning with dual label correlation perspectives. In *Proceedings of the International Joint Conference on Artificial Intelligence*, pages 4479–4487, 2024.

[Liu *et al.*, 2021] Weiwei Liu, Haobo Wang, Xiaobo Shen, and Ivor W Tsang. The emerging trends of multi-label learning. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(11):7955–7974, 2021.

[Liu *et al.*, 2023] Bing-Qing Liu, Bin-Bin Jia, and Min-Ling Zhang. Towards enabling binary decomposition for partial multi-label learning. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2023.

[Liu *et al.*, 2024] Chengliang Liu, Gehui Xu, Jie Wen, Yabo Liu, Chao Huang, and Yong Xu. Partial multi-view multi-label classification via semantic invariance learning and prototype modeling. In *Proceedings of the International Conference on Machine Learning*, 2024.

[Lyu *et al.*, 2020] Gengyu Lyu, Songhe Feng, and Yidong Li. Partial multi-label learning via probabilistic graph matching mechanism. In *Proceedings of the ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pages 105–113, 2020.

[Lyu *et al.*, 2021] Gengyu Lyu, Songhe Feng, and Yidong Li. Noisy label tolerance: A new perspective of partial multi-label learning. *Information Science*, 543:454–466, 2021.

[Lyu *et al.*, 2023] Gengyu Lyu, Songhe Feng, Yi Jin, Tao Wang, Congyan Lang, and Yidong Li. Prior knowledge regularized self-representation model for partial multilabel learning. *IEEE Transactions on Cybernetics*, 53(3):1618–1628, 2023.

[Maddison *et al.*, 2017] Chris J Maddison, Andriy Mnih, and Yee Whye Teh. The concrete distribution: A continuous relaxation of discrete random variables. In *Proceedings of the International Conference on Learning Representations*, 2017.

[Read *et al.*, 2011] Jesse Read, Bernhard Pfahringer, Geoff Holmes, and Eibe Frank. Classifier chains for multi-label classification. *Machine learning*, 85(3):333–359, 2011.

[Sun *et al.*, 2019] Lijuan Sun, Songhe Feng, Tao Wang, Congyan Lang, and Yi Jin. Partial multi-label learning with low-rank and sparse decomposition. In *Proceedings of the Association for the Advancement of Artificial Intelligence*, pages 5016–5023, 2019.

[Sun *et al.*, 2021] Lijuan Sun, Songhe Feng, Gengyu Lyu, Hua Zhang, and Guojun Dai. Partial multi-label learning with noisy side information. *Knowledage Information Systems*, 63(2):541–564, 2021.

[Sun *et al.*, 2022] Lijuan Sun, Songhe Feng, Jun Liu, Gengyu Lyu, and Congyan Lang. Global-local label correlation for partial multi-label learning. *IEEE Transactions on Multimedia*, 24:581–593, 2022.

[Vaswani *et al.*, 2017] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. *Advances in neural information processing systems*, 30, 2017.

[Wang *et al.*, 2019] Haobo Wang, Weiwei Liu, Yang Zhao, Chen Zhang, Tianlei Hu, and Gang Chen. Discriminative and correlative partial multi-label learning. In *Proceedings of the International Joint Conference on Artificial Intelligence*, pages 3691–3697, 2019.

[Wang *et al.*, 2023] Haobo Wang, Shisong Yang, Gengyu Lyu, Weiwei Liu, Tianlei Hu, Ke Chen, Songhe Feng, and Gang Chen. Deep partial multi-label learning with graph disambiguation. *arXiv preprint arXiv:2305.05882*, 2023.

[Wen *et al.*, 2024] Jie Wen, Chengliang Liu, Shijie Deng, Yicheng Liu, Lunke Fei, Ke Yan, and Yong Xu. Deep double incomplete multi-view multi-label learning with incomplete labels and missing views. *IEEE Transactions Neural Networks Learning System*, 35(8):11396–11408, 2024.

[Wu *et al.*, 2020] Jinghan Wu, Xuan Wu, Qingguo Chen, Yao Hu, and Minling Zhang. Feature-induced manifold disambiguation for multi-view partial multi-label learning. In *Proceedings of the ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pages 557–565, 2020.

[Xie and Huang, 2018] Ming-Kun Xie and Sheng-Jun Huang. Partial multi-label learning. In *Proceedings of the Association for the Advancement of Artificial Intelligence*, pages 4302–4309, 2018.

[Xie and Huang, 2021] Ming-Kun Xie and Sheng-Jun Huang. Partial multi-label learning with noisy label identification. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(7):3676–3687, 2021.

[Xie *et al.*, 2021] Ming-Kun Xie, Feng Sun, and Sheng-Jun Huang. Partial multi-label learning with meta disambiguation. In *Proceedings of the ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pages 1904–1912, 2021.

[Xu *et al.*, 2020] Ning Xu, Yun-Peng Liu, and Xin Geng. Partial multi-label learning with label distribution. In *Proceedings of the Association for the Advancement of Artificial Intelligence*, pages 6510–6517, 2020.

[Xu *et al.*, 2021] Ning Xu, Yun-Peng Liu, Yan Zhang, and Xin Geng. Progressive enhancement of label distributions for partial multilabel learning. *IEEE Transactions on Neural Networks and Learning Systems*, 2021.

[Xu *et al.*, 2022] Jiazhi Xu, Shen Huang, Fengtao Zhou, Luwen Huangfu, Daniel Zeng, and Bo Liu. Boosting multi-label image classification with complementary parallel self-distillation. In *Proceedings of the International Joint Conference on Artificial Intelligence*, pages 1495–1501, 2022.

[Xu *et al.*, 2024] Tiantian Xu, Yuanyuan Xu, Shiyu Yang, Binghao Li, and Wenjie Zhang. Learning accurate label-specific features from partially multilabeled data. *IEEE Transactions on Neural Networks and Learning Systems*, 35(8):10436–10450, 2024.

[Yan and Guo, 2021] Yan Yan and Yuhong Guo. Adversarial partial multi-label learning with label disambiguation. In *Proceedings of the Association for the Advancement of Artificial Intelligence*, pages 10568–10576, 2021.

[Yan *et al.*, 2021] Yan Yan, Shining Li, and Lei Feng. Partial multi-label learning with mutual teaching. *Knowledge-Based Systems*, 212:106624, 2021.

[Yu *et al.*, 2018] Guoxian Yu, Xia Chen, Carlotta Domeniconi, Jun Wang, Zhao Li, Zili Zhang, and Xindong Wu. Feature-induced partial multi-label learning. In *Proceedings of the IEEE International Conference on Data Mining*, pages 1398–1403, 2018.

[Zhang and Fang, 2021] Min-Ling Zhang and Jun-Peng Fang. Partial multi-label learning via credible label elicitation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 43(10):3587–3599, 2021.

[Zhang and Wu, 2015] Min-Ling Zhang and Lei Wu. LIFT: Multi-label learning with label-specific features. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 37(1):107–120, 2015.

[Zhang and Zhang, 2010] Min-Ling Zhang and Kun Zhang. Multi-label learning by exploiting label dependency. In *Proceedings of the ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pages 999–1008, 2010.

[Zhang and Zhou, 2014] Min-Ling Zhang and Zhi-Hua Zhou. A review on multi-label learning algorithms. *IEEE Transactions on Knowledge and Data Engineering*, 26(8):1819–1837, 2014.

[Zhao *et al.*, 2022] Peng Zhao, Shiyi Zhao, Xuyang Zhao, Huiting Liu, and Xia Ji. Partial multi-label learning based on sparse asymmetric label correlations. *Knowledge-Based Systems*, 245:108601, 2022.