Pseudo-Label Reconstruction for Partial Multi-Label Learning

Yu Chen[1] , **Fang Li**[1] , **Na Han** [2] , **Guanbin Li** [3] ,
Hongbo Gao[4] , **Sixian Chan**[5] and **Xiaozhao Fang**[1*]
[1]Guangdong University of Technology
[2]Guangdong Polytechnic Normal University
[3]Sun Yat-sen University
[4]Institute of Advanced Technology, University of Science and Technology of China
[5]Zhejiang University Of Technology
{chenyu9265324, lifang430481}@163.com, {hannagdut, xzhfang168}@126.com,
liguanbin@mail.sysu.edu.cn, ghb48@ustc.edu.cn, sxchan@zjut.edu.cn

Abstract

In Partial Multi-Label Learning (PML), each instance is associated with a candidate label set containing multiple relevant labels along with other false positive labels. Currently, most PML methods directly extract instance correlation from instance features while ignoring the candidate labels, which may contain more discriminative instance-related information. This paper argues that, with a well-designed model, more accurate instance correlation can be mined from the candidate labels to facilitate label disambiguation. To this end, we propose a novel PML method based on pseudo-label reconstruction (PML-PLR). Specifically, we first propose a novel orthogonal candidate label reconstruction method, which jointly optimizes with instance features to extract more consistent instance correlation. Then, we use instance correlation as reconstruction coefficient to reconstruct pseudo-labels. Subsequently, through local manifold learning, the reconstructed pseudo-labels are leveraged to propagate the consistency relationship between labels and instances, thereby improving the accuracy of pseudo-labels. Extensive experiments and analyses demonstrate that the proposed PML-PLR outperforms state-of-the-art methods.

1 Introduction

In the domain of machine learning, entities from the real world are often encapsulated as data points, each comprised of features and a corresponding label. In conventional classification tasks, labels are exclusive, allowing only a single label to be assigned to each data point. However, numerous real-world items possess multiple layers of meaning. Consider a film that might encompass genres like science fiction, war, and adventure, or a news piece that could be categorized under politics, economics, and sports. Multi-label learning (MLL) [Liu *et al.*, 2021], which permits the assignment of multiple discrete labels to a single data point, has thus gained significant attention in machine learning circles and is utilized in a range of practical applications [Tahzeeb and Hasan, 2022], including: text classification, image recognition and protein function prediction. However, the process of acquiring datasets with accurate labels is not only costly but also difficult. More often, we may end up with just a candidate set of labels, some of which are relevant and others are just noise. This noise will blur the classification decision boundary, reducing the effectiveness of MLL.

To address this challenge, Xie and Huang [Xie and Huang, 2018] propose the concept of partial multi-label learning (PML) as an innovative framework, whose objective is to build a model capable of assigning labels to new instances with noisy labels. The main challenge in PML is detecting noisy labels and building accurate classification models, which traditional multi-label learning (MLL) algorithms fail to address. For example, ML-KNN [Zhang and Zhou, 2007] and LIFT [Zhang and Wu, 2014] cannot effectively handle noisy labels, leading to poor performance on noisy datasets.

In recent years, the development of PML methods can be broadly categorized into two types: two-stage methods and end-to-end methods [Chen *et al.*, 2024]. In two-stage methods, noisy labels are first identified and then the classifier is trained. Such as, DRAMA [Wang *et al.*, 2019] constructs a label confidence matrix based on the feature manifold and candidate label set, utilizing it directly to train a multi-output regression model. PARTICLE [Zhang and Fang, 2020] applies label propagation to integrate information from K-nearest neighbors, establishes a confidence threshold for labels, and subsequently trains the MLC classifier. In contrast, end-to-end methods optimize reliable labels and the classifier simultaneously. Examples include PML-DNDC [Hu *et al.*, 2023] employs dual noise elimination, simultaneously removing label and feature noise to enhance classifier training by exploring label dependencies and promoting label similarity among classifiers. PML-ND [Zhong *et al.*, 2024] employs negative label information to guide label propagation process to induce ground-truth labels with high credibility.

These methods address PML problems, but several critical limitations remain: 1) Existing approaches largely overlook

*Corresponding author

the discriminative instance correlations embedded within candidate labels, primarily because noise contamination prevents direct extraction of meaningful correlations between instances. 2) Current disambiguation methods produce pseudo-labels that inadequately capture the intrinsic consistency relationship between feature representations and label semantics, leading to suboptimal classification performance.

To address these limitations, this paper proposes a novel PML method based on pseudo-label reconstruction (PML-PLR). Specifically, we first introduce an innovative orthogonal candidate label reconstruction approach that leverages subspace reconstruction techniques to effectively mitigate both noise interference and complex label correlations within candidate label sets. This reconstruction is then jointly optimized with instance features to extract more robust and semantically meaningful instance correlations that better reflect true label relationships. Subsequently, these refined instance correlations serve as reconstruction coefficients to generate high-quality pseudo-labels that better approximate ground-truth labels. Finally, through local manifold learning, the reconstructed pseudo-labels propagate consistency relationships between features and labels, ensuring the final pseudo-labels align more closely with the underlying true label distribution. The main framework is illustrated in Figure 1.

2 Related Work

2.1 Multi-Label Learning

In multi-label learning (MLL), each example is associated with multiple valid labels, and various approaches are extensively studied. Some methods transform MLL into binary classification problems, treating each label independently [Zhang *et al.*, 2024]. To enhance performance, many studies explore label correlation, including pairwise and high-order dependencies. Recently, integrating manifold learning with MLL has gained increasing attention. Hou et al. [Hou *et al.*, 2016] investigate the manifold structure in the label space, assuming that correlated instances share labels. Zhang et al. [Zhang *et al.*, 2019] employ manifold learning and sparse feature selection to obtain a low-dimensional embedding for label information. Zhao et al. [Zhao *et al.*, 2022b] combine manifold and subspace learning to mitigate noise and handle missing labels, reconstructing a more robust feature and label space. However, MLL typically assumes that each instance is precisely labeled, which is often unrealistic [Wang *et al.*, 2024; Wang *et al.*, 2023; Wang *et al.*, 2021; Wen *et al.*, 2023; Wen *et al.*, 2022; Wen *et al.*, 2018]. In noisy MLL, binary labels may be flipped to incorrect values. While these approaches reduce the burden of multi-label annotation, they neglect the inherent challenge of labeling—specifically, that labels themselves can be ambiguous.

2.2 Partial Multi-Label Learning

Compared to MLL, PML presents greater challenges, requiring learning in imperfect environments and training high-precision classifiers. Current PML research primarily centers on label disambiguation—identifying true labels within candidate sets. Existing approaches typically employ explicit learning strategies to remove noisy labels through various techniques. Some methods leverage low-rank and sparse decomposition to separate ground-truth labels from noise [Sun *et al.*, 2019; Sun *et al.*, 2024], while others estimate candidate label credibility via label propagation or enhancement techniques [Zhang and Fang, 2020; Xu *et al.*, 2020; Xu *et al.*, 2023; Lin *et al.*, 2025; Chen *et al.*, 2024]. Many approaches utilize feature information to identify noise [Yu *et al.*, 2018; Xie and Huang, 2021; Wang *et al.*, 2019; Lyu *et al.*, 2021], with the most common practice being the application of correlation or manifold constraints on labels [Li *et al.*, 2021; Lyu *et al.*, 2020; Li *et al.*, 2022; Hu *et al.*, 2023]. Feature selection has recently emerged as a popular disambiguation strategy [Wang *et al.*, 2022; Hao *et al.*, 2023; Han *et al.*, 2025; Wu *et al.*, 2025], while other researchers harness label correlations or cluster assignments [Zhao *et al.*, 2022a; Sun *et al.*, 2021; Qian *et al.*, 2024a; Wang *et al.*, 2025; Fang *et al.*, 2025]. Various deep learning algorithms have also been deployed to mine data distributions and mitigate noise impact [Hang and Zhang, 2023; Xie and Huang, 2022; Qian *et al.*, 2024b].

3 Proposed Method

In this section, we introduce PML-PLR and its feasibility optimization. Define some common variables as: $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_n] \in \mathbb{R}^{d \times n}$ denotes feature matrix of n instances with d-dimensional features. $\mathbf{Y} \in \{0,1\}^{l \times n}$ represents a candidate label matrix with noisy information, where l is the number of label classes. If $\mathbf{Y}_{ij} = 1$ then the i-th label is associated with the j-th instance. If $\mathbf{Y}_{ij} = 0$, the opposite. Each instance correspond to a set of candidate labels with unrelated labels incorrectly labeled as 1, which is called noisy label. The goal of PML is to minimize the impact of noisy information to make correct label predictions.

3.1 Learning Objective

Our ultimate goal is to induce a multi-label predictor $f : \mathbf{X} \mapsto [0,1]^l$, which can assign an appropriate set of labels to unseen instances. The main objectives of this paper are as follows:

1) **Learning more consistent instance-level correlation**, which allows the model to obtain more accurate correlation between instances.

2) **Learning label-instance relationships** for pseudo-label reconstruction, which makes the generated pseudo-labels more consistent with instances and the real label distribution.

3.2 Instance-level Consistent Correlation Learning

The correlation between learning instances is typically measured using the feature similarity between each pair of instances. Common methods include the Gaussian kernel function, cosine similarity, and Pearson correlation coefficient. In order to learn instance-level correlation with higher consistency between features and labels, we first use the distance relationships of each instance's features to preliminarily learn the instance-level correlation. The formula is :

$$\min_{\mathbf{C}} \sum_{i=1}^{n} \sum_{j=1}^{n} \|\mathbf{x}_i - \mathbf{x}_j\|_2^2 c_{ij}, \ \ s.t. \ \mathbf{C} \geq 0, \ \mathbf{1C} = \mathbf{1}, \quad (1)$$

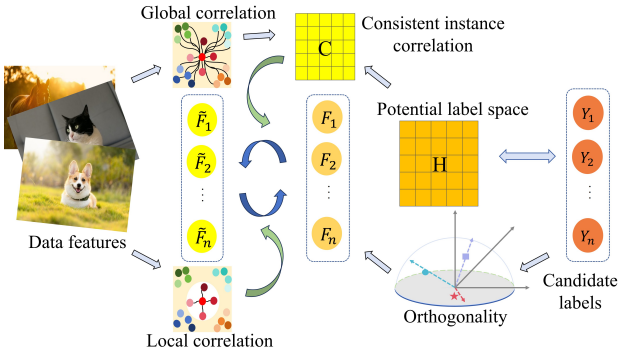


Figure 1: The main framework of PML-PLR.

where, $\mathbf{1}$ is an n-dimensional all-1 row vector, $\mathbf{C} \in \mathbb{R}^{n\times n}$ represents the captured instance correlation matrix, each element c_{ij} reflects the influence of j-th instance on i-th instance. To enhance the consistency of instance-level correlation, we choose to use label information to jointly learn correlation. We believe that both features and labels are used to describe instances, and label information contains more discriminative instance correlation. We use label self-mapping to capture the correlation between instances without losing this discriminability. The formula is expressed as follows:

$$\min_{\mathbf{C}} \sum_{i,j=1}^{n} \|\mathbf{x}_i - \mathbf{x}_j\|_2^2 c_{ij} + \|\mathbf{Y} - \mathbf{YC}\|_F^2,\ s.t. \mathbf{C} \geq 0,\ \mathbf{1C} = \mathbf{1}. \tag{2}$$

However, since label $\mathbf{Y}$ contains noise, $\mathbf{C}$ learned by the second item in the above method is not accurate enough. And due to the correlation between labels, some labels are highly related, leading to situations where two labels almost always appear together. To reduce the influence of label correlation and noise on instance-level correlation learning, for the second item, we select a more representative latent label space to learn more robust instance-level correlation. The formula is:

$$\begin{aligned} &\min_{\mathbf{P},\mathbf{H},\mathbf{C}} \|\mathbf{Y} - \mathbf{P}^\top \mathbf{H}\|_F^2 + \|\mathbf{H} - \mathbf{HC}\|_F^2, \\ &s.t.\ \ \mathbf{PP}^\top = \mathbf{I}_m,\ \mathbf{C} \geq 0,\ \mathbf{1C} = \mathbf{1}, \end{aligned} \tag{3}$$

where $\mathbf{H} \in \mathbb{R}^{m\times n} (m \leq l)$ represents the potential label space matrix, and $\mathbf{P} \in \mathbb{R}^{m\times l}$ is the indication matrix mapping the original label matrix to the latent label space. The constraint $\mathbf{PP}^\top = \mathbf{I}_m$ ensures no distortion during dimensionality reduction, with the row vectors of $\mathbf{P}$ forming an orthogonal basis, meaning the column vectors of $\mathbf{H}$ are orthogonal components of $\mathbf{Y}$ in m-dimensional space. To mitigate information loss and dimensional sensitivity, we introduce a regularization term to preserve information flow. Specifically, we add a reconstruction term to retain key information from the original label matrix and a ℓ_1-norm sparsity constraint to prevent overfitting due to noisy labels. The updated formula is as follows:

$$\begin{aligned} &\min_{\mathbf{P},\mathbf{H},\mathbf{C}} \|\mathbf{Y} - \mathbf{P}^\top \mathbf{HC}\|_F^2 + \alpha \|\mathbf{Y} - \mathbf{P}^\top \mathbf{H}\|_1, \\ &s.t.\ \ \mathbf{PP}^\top = \mathbf{I}_m,\ \mathbf{C} \geq 0,\ \mathbf{1C} = \mathbf{1}, \end{aligned} \tag{4}$$

where α is a regularization parameter controlling the importance of preserving the original label structure. By updating Eq.(2), the learning formula of the final instance-level correlation is as follows:

$$\begin{aligned} \min_{\mathbf{P},\mathbf{H},\mathbf{C}} &\sum_{i,j=1}^{n} \|\mathbf{x}_i - \mathbf{x}_j\|_2^2 c_{ij} + \|\mathbf{Y} - \mathbf{P}^\top \mathbf{HC}\|_F^2 \\ &+ \alpha \|\mathbf{Y} - \mathbf{P}^\top \mathbf{H}\|_1, \\ s.t.\ \ &\mathbf{PP}^\top = \mathbf{I}_m,\ \mathbf{C} \geq 0,\ \mathbf{1C} = \mathbf{1}. \end{aligned} \tag{5}$$

3.3 Pseudo-label Reconstruction

In order to obtain better classification prediction results, pseudo-label learning should maintain high consistency between instances and labels while minimizing noise. Furthermore, to maximize the effectiveness of previously obtained instance-level correlation in subsequent pseudo-label reconstruction, we need to reduce the influence of incorrect instance correlations present in the original labels. We achieve this by utilizing a rotation transformation to alter the instance distribution, thereby weakening the instance correlations embedded in the original labels. This rotation-based approach is formulated as follows:

$$\min_{\mathbf{F},\mathbf{Q}} \|\mathbf{YQ} - \mathbf{F}\|_F^2, s.t.\ \ \mathbf{Q}^\top \mathbf{Q} = \mathbf{I}_n,\ \mathbf{0}_{l\times n} \leq \mathbf{F} \leq \mathbf{Y}, \tag{6}$$

where $\mathbf{F} = [\mathbf{f}_1, \mathbf{f}_2, \cdots, \mathbf{f}_n] \in \mathbb{R}^{l\times n}$ represents the pseudo-label matrix and $\mathbf{Q} \in \mathbb{R}^{n\times n}$ represents the orthogonal projection transformation matrix. By constraining $\mathbf{Q}$ to be orthogonal, the resulting pseudo-label matrix $\mathbf{F}$ preserves the essential structure while minimizing the misleading correlations inherited from the original label matrix $\mathbf{Y}$. The constraint of $\mathbf{0}_{l\times n} \leq \mathbf{F} \leq \mathbf{Y}$ can ensure that the main structure of the pseudo-label is consistent with the original label.

Next, we use the previously learned instance-level correlation $\mathbf{C}$ as the reconstruction coefficient to reconstruct the pseudo-labels, endowing them with the correct instance correlations. This approach follows the principle that similar instances should have similar labels. For the i-th instance $\mathbf{f}_i$, the instance-level correlation with other instances is used as label weight to reconstruct it. We can obtain reconstructed pseudo-label $\tilde{\mathbf{f}}_i$:

$$\tilde{\mathbf{f}}_i = \frac{\sum_{j=1}^{n} (\mathbf{f}_j * c_{ij})}{\sum_{i=1}^{n} c_{ij}} = \sum_{j=1}^{n} (\mathbf{f}_j * c_{ij}) = \mathbf{F}\mathbf{c}_i. \tag{7}$$

To enhance the consistency between pseudo-labels and instances, we employ nearest neighbor relationships to constrain distances between pseudo-labels and their reconstructed counterparts. This approach ensures labels conform to the data's local manifold structure while propagating label-instance relationships. The neighbor correlation is computed:

$$\mathbf{K}_{ij} = \begin{cases} exp(\frac{-\|\mathbf{x}_i - \mathbf{x}_j\|_2^2}{2\sigma^2}), & \mathbf{x}_j \in \mathcal{N}(\mathbf{x}_j)\ or\ \mathbf{x}_j \in \mathcal{N}(\mathbf{x}_i), \\ 0, & otherwise, \end{cases} \tag{8}$$

where $\mathcal{N}(\mathbf{x}_i)$ denotes the set of k nearest neighbors of sample $\mathbf{x}_i$, and $\sigma = \frac{1}{n}\sum_{i=1}^{n} \|\mathbf{x}_i - \mathbf{x}_i^{(k)}\|_2$ is the bandwidth parameter

of the Gaussian kernel, which controls how the distance between samples affects their similarity. The main goal of this section is to first weaken the instance-level correlation in the original labels using orthogonal mapping, then use the previously learned, more consistent instance-level correlation to reconstruct the pseudo-labels, and apply local manifold constraints to restore the instance-level correlation of the pseudo-labels. The relationship between the label and the instance is propagated by reconstructing the pseudo-label. The overall formula for this section is expressed as follows:

$$\begin{aligned} &\min_{\mathbf{F},\mathbf{Q}} \|\mathbf{YQ}-\mathbf{F}\|_F^2+\beta\sum_{i=1}^{n}\sum_{j=1}^{n}\|\mathbf{f}_i-\mathbf{F}\mathbf{c}_j\|_2^2\mathbf{K}_{ij}, \\ &s.t.\ \ \mathbf{Q}^\top\mathbf{Q}=\mathbf{I}_n,\ \mathbf{0}_{l\times n}\le\mathbf{F}\le\mathbf{Y}. \end{aligned} \tag{9}$$

3.4 Kernel Nonlinear Classifier

For the classifier part, we represent $\phi(\bullet):\mathbb{R}^d\to\mathbb{R}^h$ as a feature mapping from the feature space to some h-dimensional high-dimensional Hilbert space, and then we can train the ridge regression model. Further using the deviation term $\mathbf{b}\in\mathbb{R}^c$ in the prediction function, the prediction function with kernel extension can be expressed as $\mathbf{g}(\mathbf{x}_i)=\mathbf{W}^\top\phi(\boldsymbol{x}_i)+\mathbf{b}$. Supervised training using the pseudo-label $\mathbf{F}$, then the classification model can be expressed as:

$$\min_{\mathbf{W}}\sum_i\|\boldsymbol{\epsilon}_i\|_2^2+\lambda\|\mathbf{W}\|_{\mathrm{F}}^2,\ \ s.t.\ \boldsymbol{f}_i=\mathbf{W}^\top\phi(\boldsymbol{x}_i)+\mathbf{b}-\boldsymbol{\epsilon}_i \tag{10}$$

By defining a matrix $\mathbf{E}=[\boldsymbol{\epsilon}_1,\boldsymbol{\epsilon}_2,\ldots,\boldsymbol{\epsilon}_3]^\top\in\mathbb{R}^{l\times n}$ the above problem can be re-written in the following matrix form:

$$\min_{\mathbf{W}} tr(\mathbf{E}^\top\mathbf{E})+\lambda\, tr(\mathbf{W}^\top\mathbf{W}),\ s.t.\ \mathbf{F}=\mathbf{W}^\top\mathbf{\Phi}+\mathbf{b1}-\mathbf{E}, \tag{11}$$

where $\mathbf{\Phi}=[\phi(x_1),\phi(x_2),\ldots,\phi(x_n)]\in\mathbb{R}^{h\times n}$. To facilitate optimization, we restate Eq.(11) as:

$$\min_{\mathbf{W},\mathbf{b}}\|\mathbf{W}^\top\mathbf{\Phi}+\mathbf{b1}-\mathbf{F}\|_{\mathbf{F}}^2+\lambda\{\|\mathbf{W}\|_{\mathbf{F}}^2+\|\mathbf{b}\|_{\mathbf{F}}^2\}. \tag{12}$$

3.5 Overall Objective Function

By combining the Eq.(5) (consistent correlation learning), Eq.(9) (pseudo-label learning and reconstruction) and Eq.(12) (classifier), the final objective function of the model is obtained as follows:

$$\begin{aligned} &\min_{\mathbf{W},\mathbf{b},\mathbf{P},\mathbf{H},\mathbf{C},\mathbf{Q},\mathbf{F}}\|\mathbf{W}^\top\Phi+\mathbf{b1}-\mathbf{F}\|_F^2+\|\mathbf{YQ}-\mathbf{F}\|_F^2+ \\ &\sum_{i,j=1}^{n}\|\mathbf{x}_i-\mathbf{x}_j\|_2^2c_{ij}+\|\mathbf{Y}-\mathbf{P}^\top\mathbf{HC}\|_F^2+\alpha\|\mathbf{Y}-\mathbf{P}^\top\mathbf{H}\|_1 \\ &+\beta\sum_{i,j=1}^{n}\|\mathbf{f}_i-\mathbf{F}\mathbf{c}_j\|_2^2\mathbf{K}_{ij}+\lambda\{\|\mathbf{W}\|_F^2+\|\mathbf{b}\|_F^2\}, \\ &s.t.\ \ \mathbf{Q}^\top\mathbf{Q}=\mathbf{I}_n,\ \ \mathbf{C}\ge 0,\ \mathbf{1C}=\mathbf{1}, \\ &\qquad\mathbf{PP}^\top=\mathbf{I}_m,\ \ \mathbf{0}_{l\times n}\le\mathbf{F}\le\mathbf{Y}. \end{aligned} \tag{13}$$

3.6 Optimization

For the above optimization problem, ADMM [Feng *et al.*, 2020] is adopted to solve it. In other words, we use alternate optimization to solve Eq.(13), details are as follows:

Update P and H

Removing the items that are irrelevant to $\mathbf{P}$ and $\mathbf{H}$, and fix $\mathbf{C}$. Then, the sub-optimization as:

$$\begin{aligned} &\min_{\mathbf{P},\mathbf{H},\mathbf{E}}\|\mathbf{Y}-\mathbf{P}^\top\mathbf{HC}\|_F^2+\alpha\|\mathbf{E}\|_1, \\ &s.t.\ \ \mathbf{PP}^\top=\mathbf{I}_m,\ \mathbf{Y}-\mathbf{P}^\top\mathbf{H}=\mathbf{E}. \end{aligned} \tag{14}$$

Then we apply the augmented Lagrange multiplier method to get the following Lagrange function:

$$\begin{aligned} &\min_{\mathbf{P},\mathbf{H},\mathbf{E}}\|\mathbf{Y}-\mathbf{P}^\top\mathbf{HC}\|_F^2+\alpha\|\mathbf{E}\|_1+\frac{\mu}{2}\|\mathbf{Y}-\mathbf{P}^\top\mathbf{H}-\mathbf{E}\|_F^2, \\ &s.t.\ \ \mathbf{PP}^\top=\mathbf{I}_m. \end{aligned} \tag{15}$$

where μ is a very large number. With $\mathbf{P}$ and $\mathbf{E}$ fixed, since $\mathbf{PP}^\top=\mathbf{I}_m$, $\mathbf{H}$ can be solved in a closed form :

$$\mathbf{H}^{(t+1)}=(2\mathbf{PYC}^\top-u\mathbf{P}(\mathbf{Y}-\mathbf{E}))(2\mathbf{CC}^\top-\mu\mathbf{I}_n)^{-1}. \tag{16}$$

With $\mathbf{H}$ and $\mathbf{E}$ fixed, for $\mathbf{P}$, the sub-optimization can be converted to:

$$\max_{\mathbf{P}}\ Tr((2\mathbf{YC}^\top\mathbf{H}^\top+\mu(\mathbf{Y}-\mathbf{E})\mathbf{H}^\top)\mathbf{P}),\ s.t.\ \mathbf{PP}^\top=\mathbf{I}_m, \tag{17}$$

where $Tr(\cdot)$ is the trace norm. It is noted that the problem in Eq.(17) corresponds to the classic orthogonal procrustes problem [Cai *et al.*, 2010], which can be approximated by singular value decomposition (SVD). Let $\mathbf{R}=2\mathbf{YC}^\top\mathbf{H}^\top+u(\mathbf{Y}-\mathbf{E})\mathbf{H}^\top$, then we utilize SVD to decompose $\mathbf{R}$, i.e., $\mathbf{R}=\mathbf{U\Sigma V}^\top$, optimization of $\mathbf{P}$ can be obtained by:

$$\mathbf{P}^{(t+1)}=\mathbf{VU}^\top. \tag{18}$$

With $\mathbf{H}$ and $\mathbf{P}$ fixed, the variables $\mathbf{E}$ can be optimized by solving following problem:

$$\min_{\mathbf{E}}\ \mathcal{L}(\mathbf{E})=\alpha\|\mathbf{E}\|_1+\frac{\mu}{2}\|\mathbf{Y}-\mathbf{P}^\top\mathbf{H}-\mathbf{E}\|_F^2, \tag{19}$$

which is a typical LASSO regression problem [Mirone and Paleo, 2017], and we apply PGD algorithm to optimize it. The proximal operator of Eq.(19) is:

$$\mathrm{prox}_{th(\cdot)}(\mathbf{E})=arg\min_{\mathbf{E}}\|\mathbf{E}-\mathbf{Z}\|_F^2+\frac{\alpha}{\mu\mathbf{L}}\|\mathbf{E}\|_1\ , \tag{20}$$

where $\mathbf{Z}=\mathbf{E}^t-\frac{1}{\mathbf{L}}\nabla\mathcal{L}(\mathbf{E}^{(t)})$, $\mathbf{E}^{(t)}$ represents the solution from the t-th iteration, and $\nabla\mathcal{L}(\mathbf{E})$ is the gradient of the objective function $\mathcal{L}(\mathbf{E})$, $\mathbf{L}$ is the Lipschitz constant of $\nabla\mathcal{L}(\mathbf{E})$ and t denotes the number of iteration. Problem of Eq.(20) can be iteratively updated by the soft-thresholding operator:

$$\mathbf{E}_{ij}^{(t+1)}=Soft[\mathbf{E}_{ij}^{(t)}-\frac{1}{\mathbf{L}}\nabla\mathcal{L}(\mathbf{E}_{ij}^t,\frac{\alpha}{\mu\mathbf{L}}], \tag{21}$$

where $Soft[b,\nu]=sign(b)max\{|b|-\nu,0\}$. In addition, the Lipschitz constant of $\nabla\mathcal{L}(\mathbf{E})$ is 1, so we set $\mathbf{L}=1$.

Update C

Removing the items that are irrelevant to $\mathbf{C}$, the suboptimization for $\mathbf{C}$ is simplified as:

$$\min_{\mathbf{C}}\ Tr(\mathbf{GC}^\top)+\|\mathbf{Y}-\mathbf{P}^\top\mathbf{HC}\|_F^2+2\beta Tr(\mathbf{FL}_k\mathbf{C}^\top\mathbf{F}^\top), \tag{22}$$

where $\mathbf{L}_k$ is the Laplacian matrix of $\mathbf{K}$, $\mathbf{G} \in \mathbb{R}^{n\times n}$ represents the distance matrix, for each element $\mathbf{G}_{ij} = \|\mathbf{x}_i - \mathbf{x}_j\|_2^2$. Taking the derivative of Eq.(22) and setting the derivative to zero, we obtain the following equation:

$$\mathbf{G} + 2\mathbf{H}^\top\mathbf{P}\mathbf{P}^\top\mathbf{H}\mathbf{C} - 2\mathbf{H}^\top\mathbf{P}\mathbf{Y} - 2\beta\mathbf{F}^\top\mathbf{F}\mathbf{L}_k = 0. \quad (23)$$

Then, we can obtain the optimization result variable for the first step of optimizing $\mathbf{C}$:

$$\mathbf{O} = (2\mathbf{H}^\top\mathbf{P}\mathbf{P}^\top\mathbf{H})^{-1}(2\mathbf{H}^\top\mathbf{P}\mathbf{Y} + 2\beta\mathbf{F}^\top\mathbf{F}\mathbf{L}_k - \mathbf{G}). \quad (24)$$

Then the optimization of $\mathbf{C}$ can be transformed into:

$$\min_{\mathbf{c}_i} \frac{1}{2}\|\mathbf{c}_i - \mathbf{o}_i\|_2^2, \ \ s.t\ \mathbf{c}_i \geq 0,\ \mathbf{1}\mathbf{c}_i = 1. \quad (25)$$

Eq (25) can be solved using the technique reported in [Huang *et al.*, 2015] or be solved via off-the-shelf QP tools.

Update Q

For $\mathbf{Q}$ optimization, which also is a classical orthogonal problem, $\mathbf{Q}$ can be solved as: first compute the singular-value decomposition (SVD) of matrix $\mathbf{Y}^\top\mathbf{F}$ as $\mathbf{Y}^\top\mathbf{F} = \mathbf{M}\mathbf{\Sigma}\mathbf{N}^\top$ and then let $\mathbf{Q} = \mathbf{M}\mathbf{N}^\top$.

Update W and b

We add a bias term $\mathbf{b}$ to the classifier, so that classifier $\mathbf{W}$ becomes: $\mathbf{W_m} = [\mathbf{W}; \mathbf{b}]$: the feature matrix $\mathbf{\Phi}$ becomes: $\mathbf{\Phi_m} = [\mathbf{\Phi}; \mathbf{1}_h]$. The optimization can be reformulated as:

$$\min \mathbf{W_m}\|\mathbf{W_m}^\top\mathbf{\Phi_m} - \mathbf{F}\|_F^2 + \lambda\|\mathbf{W_m}\|_F^2. \quad (26)$$

After taking the derivation and making the derivation result zero, the updated formula of $\mathbf{W_m}$ is simplified as follows:

$$\mathbf{W_m} = (\mathbf{\Phi_m}\mathbf{\Phi_m}^\top + \lambda\mathbf{I}_{h+1})^{-1}(\mathbf{X}\mathbf{F}^\top). \quad (27)$$

Update F

Removing the items that are irrelevant to $\mathbf{F}$, and fix $\mathbf{W}, \mathbf{b}, \mathbf{Q}$ and $\mathbf{C}$. Then, the suboptimization for $\mathbf{F}$ is simplified as:

$$\begin{aligned}&\min_{\mathbf{F}} \|\mathbf{W_m}^\top\mathbf{\Phi} - \mathbf{F}\|_F^2 + \|\mathbf{Y}\mathbf{Q} - \mathbf{F}\|_F^2 + \beta Tr(\mathbf{F}\mathbf{L}_k\mathbf{C}^\top\mathbf{F}^\top)\\ &s.t.\ \ \mathbf{0}_{l\times n} \leq \mathbf{F} \leq \mathbf{Y},\end{aligned} \quad (28)$$

For the $\mathbf{0}_{l\times n} \leq \mathbf{F} \leq \mathbf{Y}$ constraint, it can be decomposed into two non-negative constraints, $\mathbf{F} \geq 0$ and $\mathbf{Y} - \mathbf{F} \geq 0$. Next, the Lagrange multiplier method is applied to solve it:

$$\begin{aligned}\min_{\mathbf{F}} &\|\mathbf{W_m}^\top\mathbf{\Phi} - \mathbf{F}\|_F^2 + \|\mathbf{Y}\mathbf{Q} - \mathbf{F}\|_F^2 + 2\beta Tr(\mathbf{F}\mathbf{L}_k\mathbf{Z}^\top\mathbf{F}^\top)\\ &- Tr(\mathbf{\Omega}\mathbf{F}^\top) - Tr(\mathbf{\Theta}(\mathbf{Y} - \mathbf{F})^\top),\end{aligned} \quad (29)$$

where $\mathbf{\Omega}$ and $\mathbf{\Theta}$ represent the Lagrange multiplier. Taking the derivative of Eq.(29) and setting the derivative to zero, we obtain the following equation:

$$4\mathbf{F} - 2\mathbf{W_m}^\top\mathbf{\Phi} - 2\mathbf{Y}\mathbf{Q} + 2\beta\mathbf{F}(\mathbf{C}^\top\mathbf{L}_k + \mathbf{L}_k^\top\mathbf{C}) - \mathbf{\Omega} + \mathbf{\Theta} = 0 \quad (30)$$

Based on condition of Karush-Kuhn-Tucker (KKT), it can be given that: $\mathbf{\Omega}_{ij}\mathbf{F}_{ij} = 0$ and $\mathbf{\Theta}_{ij}(\mathbf{Y} - \mathbf{F})_{ij} = 0$ the update rules for $\mathbf{F}$ can be obtained as:

$$\mathbf{F}_{ij}^{(t+1)} = \mathbf{F}_{ij}^{(t)}\frac{\mathbf{B}_{ij}}{\mathbf{A}_{ij} + eps} - \frac{\mathbf{\Theta}_{ij}\mathbf{Y}_{ij}}{\mathbf{A}_{ij} + eps}, \quad (31)$$

where $\mathbf{A} = 2\mathbf{F} + \beta\mathbf{F}\mathbf{C}^\top\mathbf{L}_k + \beta\mathbf{F}\mathbf{L}_k^\top\mathbf{C}$, $\mathbf{B} = \mathbf{W_m}^\top\mathbf{\Phi} + \mathbf{Y}\mathbf{Q}$, eps is a tiny value to prevent the denominator from being 0. The update for $\mathbf{\Theta}$ is based on the complementary condition: $\mathbf{\Theta}_{ij} = \max(0, \mathbf{F}_{ij} - \mathbf{Y}_{ij})$.

Datsets	#Instances	#Features	#Class	avg.#CLs	avg.#GLs
Music_emotion	6833	98	11	5.29	2.42
Music_style	6839	98	10	6.04	1.44
Mirflickr	10433	100	7	3.35	1.77
YeastBP	6139	6139	127	5.93	5.54
Emotions	593	72	6	3, (4), 5	1.86
Birds	645	260	19	3, (4), 5	1.01
Medical	978	1449	45	3, (5), 7	1.25
Image	2000	294	5	2, (3), 4	1.23
Scene	2407	294	6	3, (5), 7	1.07
Yeast	2417	103	14	(9), 11	4.24
Health	5000	612	32	(7), 9	1.67
Reference	5000	793	33	(7), 9	1.17

Table 1: General information of the eight multi-label datasets and four partial multi-label learning datasets, the experimental settings in parentheses are mainly presented.

4 Experiments

4.1 Datasets

To evaluate the generalization performance of our proposed PML-PLR method, a total of 25 datasets are used for comparative study. Specifically, the experiments are conducted on 4 real-world PML datasets and 21 synthetic PML datasets generated from 8 multi-label datasets. Detailed characteristics of all datasets are summarized in Table 1. For the 3 real-world PML datasets, including $\mathbf{Music\_emotion}$, $\mathbf{Music\_style}$, $\mathbf{Mirflickr}$ and $\mathbf{YeastBP}$ [1] . For the 8 multi-label data sets, including $\mathbf{Emotions}$, $\mathbf{Birds}$, $\mathbf{Medical}$, $\mathbf{Image}$, $\mathbf{Scene}$, $\mathbf{Yeast}$[2] , $\mathbf{Health}$ and $\mathbf{Reference}$ [Zhong *et al.*, 2024], We synthesize PML datasets by randomly adding unrelated labels together with their underlying truth labels to form a candidate tag set. The noise is added by randomly introducing false positive labels. For example, in the $emotions$ dataset, the average number of ground-truth labels are 1.86 ($avg.\#GLs$), and after adding noise, the average number of positive labels become 4 ($avg.\#CLs$), resulting in an average of 2.14 noise labels per instance. **It is worth noting that** due to the page limit, most of the following experimental synthetic datasets are based on the experimental Settings in parentheses, that is, 8 synthetic datasets.

4.2 Baselines and Implementation Details

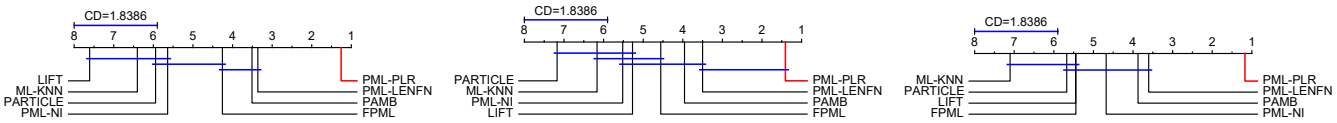
We select seven benchmark methods for comparison, including two MLL methods **ML-KNN** [Zhang and Zhou, 2007] and **LIFT** [Zhang and Wu, 2014], five state-of-the-art PML methods **PML-LENFN** [Chen *et al.*, 2024], **PAMB** [Liu *et al.*, 2023], **PML-NI** [Xie and Huang, 2021], **PARTIAL** [Zhang and Fang, 2020] and **FPML** [Yu *et al.*, 2018]. We set the corresponding parameters according to the recommendations in the respective literature. The parameters α, β and λ in the PML-PLR are selected using grid search from $\{10^{-4}, 10^{-3}, 10^{-2}, 10^{-1}, 10^{0}\}$. Cross-validation is employed to select the optimal latent space dimension m. Five widely-used multi-label metrics are employed to evaluate each comparing method, including *Hamming Loss*, *Ranking Loss*, *One-Error*, *Coverage* and *Average Precision*. Detailed definitions on them can be found in [Zhang and Zhou, 2013].

[1] http://palm.seu.edu.cn/zhangml/

[2] http://mulan.sourceforge.net/datasets.html

Datasets	PML-PLR	PML-LENFN	PAMB	PML-NI	PARTICLE	FPML	ML-KNN	LIFT
	Ranking Loss (the smaller the better)							
Music_emotion	**0.232±0.007**	0.246±0.009	0.234±0.007	0.246±0.008	0.362±0.014	0.410±0.009	0.322±0.012	0.253±0.009
Music_style	**0.132±0.012**	0.138±0.010	0.135±0.005	0.137±0.010	0.221±0.010	0.317±0.033	0.204±0.011	0.177±0.009
Mirflickr	**0.109±0.009**	0.118±0.005	0.112±0.038	0.126±0.007	0.127±0.103	0.115±0.006	0.163±0.031	0.117±0.007
YeastBP	**0.197±0.013**	0.256±0.011	0.230±0.011	0.220±0.011	0.404±0.033	0.415±0.057	0.358±0.012	0.245±0.009
	Hamming Loss (the smaller the better)							
Music_emotion	**0.208±0.004**	0.213±0.004	0.210±0.003	0.254±0.009	0.221±0.004	0.272±0.027	0.216±0.002	0.335±0.005
Music_style	**0.114±0.006**	0.116±0.005	0.115±0.004	0.159±0.012	0.125±0.004	0.312±0.048	0.123±0.005	0.819±0.006
Mirflickr	**0.169±0.004**	0.173±0.004	0.171±0.032	0.224±0.006	0.174±0.037	0.176±0.003	0.175±0.004	0.218±0.005
YeastBP	0.033±0.002	0.041±0.002	0.034±0.007	0.037±0.002	0.026±0.012	0.252±0.008	0.025±0.002	**0.024±0.002**
	Average Precision (the larger the better)							
Music_emotion	**0.632±0.011**	0.610±0.012	0.626±0.011	0.608±0.012	0.506±0.016	0.458±0.010	0.536±0.011	0.596±0.010
Music_style	**0.745±0.014**	0.731±0.016	0.741±0.007	0.739±0.015	0.657±0.012	0.566±0.090	0.674±0.015	0.681±0.015
Mirflickr	**0.825±0.013**	0.798±0.007	0.791±0.019	0.786±0.009	0.813±0.136	0.814±0.009	0.733±0.069	0.789±0.012
YeastBP	**0.416±0.021**	0.339±0.015	0.356±0.022	0.404±0.022	0.108±0.016	0.328±0.012	0.191±0.013	0.305±0.013

Table 2: Results of PML-PLR compared with other methods under real–word datasets, bold the best, underline the suboptimal.



(a) **Ranking Loss** (b) **Hamming Loss** (c) **Average Precision**

Figure 2: Results of PML-PLR against other approaches with the Nemenyi test(CD = 2.0998 at 0.05 significance level).

Evaluation metric	F_F	Critical value
Hamming Los	21.3683	
Ranking Loss	25.7514	
One Error	12.4203	2.0998
Coverage	15.0732	
Average Precision	15.1171	

Table 3: Friedman statistics F_F across five evaluation metrics and the critical value at 0.05 significance level.

4.3 Experimental Results

Due to page limitations, Tables 2 and 4 only present the results under the experimental settings in parentheses in Table 1. Table 2 reports the results of the experiment on real-world datasets. Table 4 shows the results on synthetic datasets. After analyzing the experimental results, we can draw the following conclusions:

- It turns out from Table 2 and Table 4 that we conduct 36 cases (12 datasets × 3 metrics = 36 cases) of experiments and PML-PLR performs best 31 cases (real world dataset 11 cases, synthetic dataset 20 cases), accounting for 86.11%. This proves that PML-PLR performs best on both real-world and synthetic datasets.
- MLL algorithms excel in classification on low-noise datasets, while PML algorithms perform better on high-noise datasets due to their denoising focus. PML-PLR outperforms both MLL and PML algorithms in all settings, highlighting the effectiveness of its pseudo-label reconstruction and the model's robust classification.

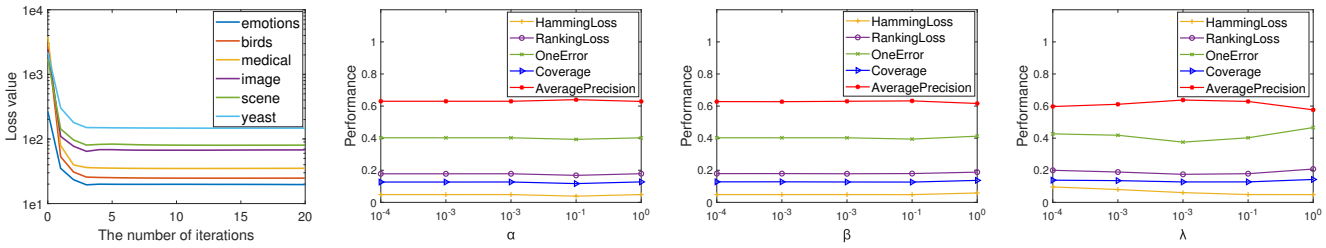
Under a total of 25 experimental settings, the Friedman test [Demšar, 2006] is used to assess the relative performance of the methods, with Table 3 showing the Friedman statistic F_F and the critical value. The post-hoc Nemenyi test evaluates statistical significance of performance differences. Figure 2 presents critical difference (CD) diagrams for each metric, where PML-PLR (highlighted with a red line) serves as the control method. Methods whose performance differences do not exceed the CD value are connected by blue lines, indicating no significant statistical difference. The results clearly highlight PML-PLR's significant superiority, as it achieves the lowest average rank across all metrics, with few or no other methods connected to it via the CD threshold.

4.4 Further Analysis

Convergence analysis: Figure 3 (a) shows the convergence curve of the model in multiple data sets. The result shows that with the increase of the number of iterations, the loss value decreases rapidly and finally stabilizes at a low value, indicating that the model converges to a satisfactory solution.
Parameter sensitivity analysis: Figure 3 (b), (c) and (d) show the sensitivity experimental results of the three parameters α, β, λ in the model. The results show that each parameter has an optimal value and PML-PLR performs stably across a wide range of parameters, making it capable of robust classification performance under various conditions.
Ablation analysis: Figure 4 shows the ablation results on two datasets, comparing PML-PLR with its three degraded versions: no pseudo-label learning (**PL-free**, which directly uses candidate labels for classifier training), no reconstruction (**R-free**, which sets parameter β=0), and no consistent

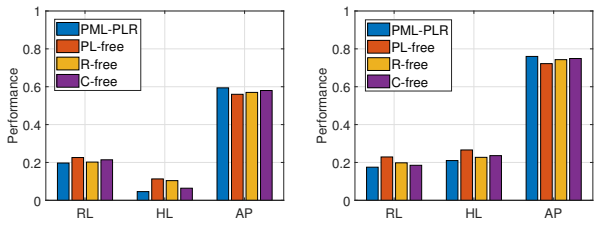
Datasets	PML-PLR	PML-LENFN	PAMB	PML-NI	PARTICLE	FPML	ML-KNN	LIFT
	Ranking Loss (the smaller the better)							
Emotions	**0.166±0.027**	0.210±0.027	0.192±0.032	0.263±0.029	0.446±0.027	0.407±0.014	0.249±0.044	0.211±0.035
Birds	**0.197±0.032**	0.200±0.028	0.204±0.024	0.205±0.034	0.326±0.027	0.341±0.020	0.250±0.041	0.214±0.030
Medical	**0.032±0.012**	0.068±0.021	0.113±0.032	0.100±0.028	0.102±0.018	0.059±0.010	0.093±0.021	0.046±0.010
Image	**0.180±0.018**	0.212±0.026	0.217±0.015	0.230±0.024	0.261±0.070	0.254±0.018	0.272±0.023	0.235±0.015
Scene	**0.134±0.011**	0.242±0.018	0.250±0.016	0.284±0.013	0.291±0.130	0.153±0.013	0.348±0.024	0.278±0.018
Yeast	**0.178±0.012**	0.190±0.016	0.211±0.007	0.202±0.016	0.189±0.009	0.191±0.017	0.193±0.015	0.188±0.011
Health	0.067±0.003	0.076±0.006	0.081±0.007	0.096±0.007	0.110±0.008	**0.063±0.003**	0.077±0.004	0.070±0.006
Reference	0.103±0.004	0.130±0.009	0.110±0.008	0.152±0.009	0.156±0.015	**0.101±0.005**	0.116±0.009	0.118±0.007
	Hamming Loss (the smaller the better)							
Emotions	**0.214±0.013**	0.247±0.022	0.221±0.022	0.500±0.047	0.253±0.022	0.330±0.013	0.261±0.022	0.570±0.035
Birds	**0.046±0.004**	0.093±0.007	0.101±0.009	0.061±0.011	0.149±0.010	0.109±0.023	0.049±0.006	0.051±0.007
Medical	0.013±0.001	0.013±0.003	0.025±0.001	0.023±0.002	0.037±0.003	0.015±0.001	0.085±0.001	**0.012±0.002**
Image	**0.203±0.009**	0.215±0.027	0.217±0.010	0.227±0.015	0.236±0.056	0.249±0.016	0.222±0.015	0.572±0.013
Scene	**0.149±0.010**	0.211±0.017	0.203±0.022	0.644±0.053	0.277±0.002	0.158±0.012	0.759±0.026	0.820±0.003
Yeast	**0.213±0.004**	0.234±0.012	0.215±0.008	0.248±0.036	0.217±0.008	0.232±0.006	0.315±0.013	0.587±0.008
Health	**0.035±0.001**	0.038±0.002	0.044±0.001	0.036±0.002	0.051±0.003	0.045±0.001	0.049±0.001	0.047±0.003
Reference	0.029±0.001	0.036±0.002	0.036±0.001	0.030±0.001	0.035±0.001	**0.027±0.001**	0.033±0.001	0.037±0.001
	Average Precision (the larger the better)							
Emotions	**0.792±0.003**	0.768±0.028	0.783±0.036	0.749±0.034	0.739±0.033	0.730±0.016	0.720±0.039	0.746±0.033
Birds	**0.594±0.032**	0.581±0.048	0.564±0.044	0.572±0.041	0.419±0.046	0.373±0.020	0.525±0.062	0.550±0.062
Medical	**0.872±0.027**	0.803±0.031	0.725±0.020	0.732±0.035	0.835±0.024	0.822±0.022	0.726±0.037	0.845±0.018
Image	**0.764±0.009**	0.747±0.029	0.748±0.019	0.732±0.024	0.725±0.084	0.696±0.023	0.689±0.022	0.720±0.019
Scene	**0.800±0.000**	0.649±0.025	0.722±0.023	0.609±0.018	0.649±0.149	0.753±0.020	0.527±0.024	0.613±0.027
Yeast	0.752±0.020	0.740±0.017	**0.753±0.013**	0.725±0.016	0.744±0.011	0.730±0.013	0.729±0.017	0.734±0.014
Health	**0.760±0.008**	0.750±0.016	0.680±0.016	0.723±0.018	0.501±0.045	0.695±0.010	0.650±0.012	0.716±0.019
Reference	**0.671±0.012**	0.643±0.012	0.612±0.015	0.614±0.011	0.448±0.058	0.576±0.014	0.589±0.007	0.619±0.014

Table 4: Results of PML-PLR compared with other methods under synthetic datasets, bold the best, underline the suboptimal.



(a) Convergence curve (b) Vary α (c) Vary β (d) Vary λ

Figure 3: Convergence curve on synthetic datasets and results of PML-PLR with varying values of trade-off parameters on Birds.



(a) Birds (b) Yeast

Figure 4: Ablation results of PML-PLR on Birds and Yeast dataset.

correlation (**C-free**, which extracts correlation only from feature matrix $\mathbf{X}$). The results clearly demonstrate that PML-PLR significantly outperforms all degraded algorithms across multiple evaluation metrics, validating the effectiveness and necessity of each component in our proposed framework. The performance drop when removing any single component highlights their complementary nature in addressing the challenges of partial multi-label learning.

5 Conclusion

This paper presents a novel PML method, PML-PLR, which jointly extracts instance-level correlation from candidate labels and features, and then uses the instance correlation as reconstruction coefficient to reconstruct pseudo-labels. Through local manifold learning, the reconstructed pseudo-labels are used to propagate the consistency relationship between labels and instances, thereby improving the accuracy of pseudo-labels. Extensive experiments demonstrate the superiority of the model.

Acknowledgments

This work was supported in part by the National Natural Science Foundation of China under Grant No. 62176065, No. 62302172, in part by the Guangdong Provincial National Science Foundation under Grant No. 2022A1515011277.

References

[Cai *et al.*, 2010] Jian-Feng Cai, Emmanuel J Candès, and Zuowei Shen. A singular value thresholding algorithm for matrix completion. *SIAM Journal on optimization*, 20(4):1956–1982, 2010.

[Chen *et al.*, 2024] Yu Chen, Yanan Wu, Na Han, Xiaozhao Fang, Bingzhi Chen, and Jie Wen. Partial multi-label learning based on near-far neighborhood label enhancement and nonlinear guidance. In *ACM Multimedia 2024*, 2024.

[Demšar, 2006] Janez Demšar. Statistical comparisons of classifiers over multiple data sets. *The Journal of Machine learning research*, 7:1–30, 2006.

[Fang *et al.*, 2025] Xiaozhao Fang, Xi Hu, Yan Hu, Yonghao Chen, Shengli Xie, and Na Han. Fuzzy bifocal disambiguation for partial multi-label learning. *Neural Networks*, 185:107137, 2025.

[Feng *et al.*, 2020] Lei Feng, Jun Huang, Senlin Shu, and Bo An. Regularized matrix factorization for multilabel learning with missing labels. *IEEE transactions on cybernetics*, 52(5):3710–3721, 2020.

[Han *et al.*, 2025] Qingqi Han, Liang Hu, and Wanfu Gao. Integrating label confidence-based feature selection for partial multi-label learning. *Pattern Recognition*, 161:111281, 2025.

[Hang and Zhang, 2023] Jun-Yi Hang and Min-Ling Zhang. Partial multi-label learning with probabilistic graphical disambiguation. *Advances in Neural Information Processing Systems*, 36:1339–1351, 2023.

[Hao *et al.*, 2023] Pingting Hao, Liang Hu, and Wanfu Gao. Partial multi-label feature selection via subspace optimization. *Information Sciences*, 648:119556, 2023.

[Hou *et al.*, 2016] Peng Hou, Xin Geng, and Min-Ling Zhang. Multi-label manifold learning. In *Proceedings of the AAAI conference on artificial intelligence*, volume 30, 2016.

[Hu *et al.*, 2023] Yan Hu, Xiaozhao Fang, Peipei Kang, Yonghao Chen, Yuting Fang, and Shengli Xie. Dual noise elimination and dynamic label correlation guided partial multi-label learning. *IEEE Transactions on Multimedia*, 2023.

[Huang *et al.*, 2015] Jin Huang, Feiping Nie, Heng Huang, et al. A new simplex sparse learning model to measure data similarity for clustering. In *IJCAI*, pages 3569–3575, 2015.

[Li *et al.*, 2021] Ziwei Li, Gengyu Lyu, and Songhe Feng. Partial multi-label learning via multi-subspace representation. In *Proceedings of the Twenty-Ninth International Conference on International Joint Conferences on Artificial Intelligence*, pages 2612–2618, 2021.

[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.

[Lin *et al.*, 2025] Yaojin Lin, Yulin Li, Shidong Lin, Lei Guo, and Yu Mao. Partial multi-label feature selection based on label distribution learning. *Pattern Recognition*, 164:111523, 2025.

[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.

[Lyu *et al.*, 2020] Gengyu Lyu, Songhe Feng, and Yidong Li. Partial multi-label learning via probabilistic graph matching mechanism. In *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & 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 Sciences*, 543:454–466, 2021.

[Mirone and Paleo, 2017] Alessandro Mirone and Pierre Paleo. A conjugate subgradient algorithm with adaptive preconditioning for the least absolute shrinkage and selection operator minimization. *Computational Mathematics and Mathematical Physics*, 57:739–748, 2017.

[Qian *et al.*, 2024a] Wenbin Qian, Yanqiang Tu, Jintao Huang, and Weiping Ding. Partial multi-label learning via robust feature selection and relevance fusion optimization. *Knowledge-Based Systems*, 286:111365, 2024.

[Qian *et al.*, 2024b] Wenbin Qian, Yanqiang Tu, Jintao Huang, Wenhao Shu, and Yiu-Ming Cheung. Partial multilabel learning using noise-tolerant broad learning system with label enhancement and dimensionality reduction. *IEEE Transactions on Neural Networks and Learning Systems*, 2024.

[Sun *et al.*, 2019] Lijuan Sun, Songhe Feng, Tao Wang, Congyan Lang, and Yi Jin. Partial multi-label learning by low-rank and sparse decomposition. In *Proceedings of the AAAI conference on artificial intelligence*, volume 33, pages 5016–5023, 2019.

[Sun *et al.*, 2021] 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, 2021.

[Sun *et al.*, 2024] Zhenzhen Sun, Zexiang Chen, Jinghua Liu, Yewang Chen, and Yuanlong Yu. Partial multi-label feature selection via low-rank and sparse factorization with manifold learning. *Knowledge-Based Systems*, 296:111899, 2024.

[Tahzeeb and Hasan, 2022] Shahab Tahzeeb and S. M. Mamun Hasan. A neural network-based multi-label classifier for protein function prediction. *Engineering, Technology & Applied Science Research*, 2022.

[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 *IJCAI*, pages 3691–3697, 2019.

[Wang *et al.*, 2021] Qianqian Wang, Zhengming Ding, Zhiqiang Tao, Quanxue Gao, and Yun Fu. Generative partial multi-view clustering with adaptive fusion and cycle consistency. *IEEE Transactions on Image Processing*, 30:1771–1783, 2021.

[Wang *et al.*, 2022] Jing Wang, Peipei Li, and Kui Yu. Partial multi-label feature selection. In *2022 International Joint Conference on Neural Networks (IJCNN)*, pages 1–9. IEEE, 2022.

[Wang *et al.*, 2023] Qianqian Wang, Zhiqiang Tao, Wei Xia, Quanxue Gao, Xiaochun Cao, and Licheng Jiao. Adversarial multiview clustering networks with adaptive fusion. *IEEE transactions on neural networks and learning systems*, 34:7635–7647, 2023.

[Wang *et al.*, 2024] Qianqian Wang, Zhiqiang Tao, Quanxue Gao, and Licheng Jiao. Multi-view subspace clustering via structured multi-pathway network. *IEEE Transactions on Neural Networks and Learning Systems*, 35(5):7244–7250, 2024.

[Wang *et al.*, 2025] Ke Wang, Yahu Guan, Yunyu Xie, Zhaohong Jia, Hong Ye, Zhangling Duan, and Dong Liang. Partial multi-label learning with label and classifier correlations. *Information Sciences*, 712:122101, 2025.

[Wen *et al.*, 2018] Jie Wen, Bob Zhang, Yong Xu, Jian Yang, and Na Han. Adaptive weighted nonnegative low-rank representation. *Pattern Recognition*, 81:326–340, 2018.

[Wen *et al.*, 2022] Jie Wen, Shijie Deng, Lunke Fei, Zheng Zhang, Bob Zhang, Zhao Zhang, and Yong Xu. Discriminative regression with adaptive graph diffusion. *IEEE Transactions on Neural Networks and Learning Systems*, 35(2):1797–1809, 2022.

[Wen *et al.*, 2023] Jie Wen, Gehui Xu, Zhanyan Tang, Wei Wang, Lunke Fei, and Yong Xu. Graph regularized and feature aware matrix factorization for robust incomplete multi-view clustering. *IEEE Transactions on Circuits and Systems for Video Technology*, 34(5):3728–3741, 2023.

[Wu *et al.*, 2025] You Wu, Peipei Li, and Yizhang Zou. Partial multi-label feature selection with feature noise. *Pattern Recognition*, 162:111310, 2025.

[Xie and Huang, 2018] Ming-Kun Xie and Sheng-Jun Huang. Partial multi-label learning. In *Proceedings of the AAAI conference on artificial intelligence*, volume 32, 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 and Huang, 2022] Ming-Kun Xie and Sheng-Jun Huang. Ccmn: A general framework for learning with class-conditional multi-label noise. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(1):154–166, 2022.

[Xu *et al.*, 2020] Ning Xu, Yun-Peng Liu, and Xin Geng. Partial multi-label learning with label distribution. In *Proceedings of the AAAI conference on artificial intelligence*, volume 34, pages 6510–6517, 2020.

[Xu *et al.*, 2023] 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*, 34(8):4856–4867, 2023.

[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 *2018 IEEE international conference on data mining (ICDM)*, pages 1398–1403. IEEE, 2018.

[Zhang and Fang, 2020] 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, 2020.

[Zhang and Wu, 2014] 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, 2014.

[Zhang and Zhou, 2007] Min-Ling Zhang and Zhi-Hua Zhou. Ml-knn: A lazy learning approach to multi-label learning. *Pattern recognition*, 40(7):2038–2048, 2007.

[Zhang and Zhou, 2013] 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, 2013.

[Zhang *et al.*, 2019] Jia Zhang, Zhiming Luo, Candong Li, Changen Zhou, and Shaozi Li. Manifold regularized discriminative feature selection for multi-label learning. *Pattern Recognition*, 95:136–150, 2019.

[Zhang *et al.*, 2024] Yao Zhang, Wei Huo, and Jun Tang. Multi-label feature selection via latent representation learning and dynamic graph constraints. *Pattern Recognition*, 151:110411, 2024.

[Zhao *et al.*, 2022a] 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.

[Zhao *et al.*, 2022b] Tianna Zhao, Yuanjian Zhang, and Witold Pedrycz. Robust multi-label classification with enhanced global and local label correlation. *Mathematics*, 10(11):1871, 2022.

[Zhong *et al.*, 2024] Jingyu Zhong, Ronghua Shang, Feng Zhao, Weitong Zhang, and Songhua Xu. Negative label and noise information guided disambiguation for partial multi-label learning. *IEEE Transactions on Multimedia*, 2024.