

Partial Multi-label learning based on shared subspace representation

Suwei Shi¹ and Rongyi Shi²

¹School of Information Science and Technology
Xiamen University Tan Kah Kee College
Zhangzhou, 363123, P.R. China
shisuwei2008@126.com

²School of Information Engineering
Zhejiang Ocean University
Zhejiang, 316002, P.R. China
shirongyi98@163.com

ABSTRACT

Partial Multi-Label Learning (PML) is a weakly supervised learning framework in which each instance is associated with a set of candidate labels, among which only a subset are correct. The presence of noisy labels in the candidate set poses significant challenges for accurate classification and effective feature selection. Existing methods often overlook either the noise in the feature space or the structural relationships within the label space, limiting their robustness and generalizability. In this paper, we propose a novel PML approach based on Shared Subspace Representation, which jointly learns a low-dimensional latent space for both feature and label information. Our method decomposes the noisy candidate label matrix into a low-rank ground-truth label component and a sparse noise component, leveraging the inherent structure and sparsity of label noise. Simultaneously, a feature correlation matrix maps original features into a compact subspace, enhancing discriminative power while mitigating the impact of feature noise. To preserve the geometric structure of the data, manifold regularization is incorporated via a graph Laplacian constructed from the feature space, enforcing local consistency between similar instances. We formulate the overall objective as a unified optimization problem, which is efficiently solved using block coordinate descent and gradient-based methods. Extensive experiments on benchmark datasets demonstrate that our approach outperforms state-of-the-art methods in both label recovery and feature selection, confirming the effectiveness of shared subspace modeling in partial multi-label learning.

Keywords: Partial multi-label learning, Label noise, Low-rank decomposition, Subspace representation, Manifold regularization..

2000 Mathematics Subject Classification: 68U35.

1 Introduction

Partial Multi-Label Learning (PML) is a weakly supervised learning paradigm that addresses scenarios where each training instance is associated with a set of candidate labels, only a subset of which are correct (Sun, Feng, Lyu, Zhang and Dai, 2021) (Xu, Liu and Geng, 2020). The key distinction between PML and related learning problems lies in the nature of the label information. In traditional Multi-Label Learning (MLL), all assigned labels are assumed to be ground truth (Feng, An and He, 2019) (Zhan and Zhang, 2017). In contrast, Partial Label Learning (PLL) assumes only one true label exists among the candidates (Lyu, Feng, Li, Jin, Dai and Lang, 2020) (Feng and An, 2019). PML generalizes these settings by allowing multiple ground-truth labels, while their exact identities remain unknown within the candidate set (Xie and Huang, 2018).

Most existing multi-label feature selection methods focus on modeling label correlations and learning mappings from features to labels (Zhang and Wu, 2014). However, in PML settings, the label information is inherently noisy and imprecise. This noise can significantly degrade the performance of traditional multi-label feature selection approaches, which often rely on accurate label supervision to identify shared discriminative features. Similarly, while PLL-based feature selection methods can be adapted to identify one ground-truth label from a candidate set, they fall short in the PML context where multiple true labels may exist and the remaining candidate labels act as noise (Xu, Wang, Shen, Wang and Chen, 2014) (Xie and Huang, 2022). A naive approach to PML is to treat all candidate labels as ground-truth and apply standard MLL techniques. However, such methods are easily misled by the inclusion of noisy labels, which can distort feature-label mappings and result in unreliable label correlations. Consequently, the selected features may not reflect the true underlying structure of the data.

To address these challenges, we propose a novel partial multi-label feature selection method that jointly learns feature mappings and decomposes noisy label information to effectively guide the model training. The key contributions of our approach are as follows:

Label Decomposition: We decompose the observed noisy label matrix into a low-rank label subspace matrix and a corresponding label correlation matrix, which enables a clearer separation between true and noisy labels.

Manifold Regularization: To preserve the intrinsic geometric structure of the feature space, we incorporate graph Laplacian regularization on the latent label subspace, ensuring that similar instances remain close in both feature and label spaces.

Feature Subspace Learning: To mitigate the impact of noise in the feature space, we introduce a feature correlation matrix that maps high-dimensional features into a low-dimensional latent subspace, improving the robustness and generalization of the model.

Joint Optimization: We formulate a unified objective that combines these components and solve it using an alternating minimization strategy based on a least squares loss function.

We formulate a unified objective that combines these components and solve it using an alternating minimization strategy based on a least squares loss function. Additionally, we employ an ℓ_1 -norm regularization strategy to separate the candidate labels into true and noisy ones, leveraging the sparsity of label noise and the low-rank structure among true labels. Extensive experiments on benchmark datasets demonstrate the effectiveness of our method. It outper-

forms existing approaches by more accurately identifying discriminative features and better handling noisy label information.

2 Related work

Several methods have been proposed to address the challenges of PML, particularly focusing on the identification and mitigation of noisy labels within the candidate label sets. PML-lc and PML-fp (Xie and Huang, 2018), proposed by Xie and Huang, are two notable approaches designed to handle label noise in PML settings. Both methods estimate the confidence score of each candidate label and incorporate these scores into the learning process by minimizing the ranking loss weighted by label confidence. In addition, they leverage the underlying data structure to refine the label confidences, thus enhancing the overall model performance. fPML (Yu, Chen, Domeniconi, Wang, Li, Zhang and Wu, 2018) further advances this direction by modeling the potential correlation between labels and features using a low-rank matrix approximation. By factorizing the original candidate label matrix, fPML is able to effectively distinguish true labels from noisy ones while simultaneously training a multi-label classifier. This decomposition allows for a more accurate representation of the latent label structure. Another significant method, PML-NI (Xie and Huang, 2022), addresses the problem by assuming that noisy labels often stem from ambiguous content in the data. PML-NI tackles the issue by decomposing the prediction model into two components: one for predicting ground-truth labels and another for identifying noisy labels. This separation enables a more targeted learning strategy that explicitly accounts for label uncertainty. MUSER introduces a robust PML framework by jointly modeling noise in both the feature space and the label space. By considering the interaction between noisy features and ambiguous labels, MUSER aims to build a model that can generalize better under real-world noisy conditions.

These methods have laid the foundation for PML research by addressing label noise and improving learning robustness. However, most of them either overlook the influence of noisy features or do not explicitly integrate feature selection with label disambiguation. In contrast, our proposed method jointly performs label decomposition and feature subspace learning, leveraging both structural consistency and sparsity to enhance model robustness and feature discriminativeness.

3 The proposed approach

In this section, we first define the PML problem addressed in this paper. We then present a detailed description of our proposed algorithm, followed by an effective optimization strategy specifically designed for our method.

3.1 Problem Formulation

Let $X = [x_1, x_2, \dots, x_n]^T \in \mathbb{R}^{n \times d}$ be the feature matrix with n instances and d features. For each instance x_i , the observed label vector is $y_i \in \{0, 1\}^q$, where q is the number of possible labels. The observed labels are potentially noisy and are stored in a candidate label matrix

$Y = [y_1, y_2, \dots, y_n]^T \in \mathbb{R}^{n \times q}$, where $Y_{ij} = 1$ represents that instance x_i is annotated with j -th label; otherwise $Y_{ij} = 0$.

3.2 Proposed methodology

We assume that the ground-truth label matrix is $\tilde{Y} = \{0, 1\}^{n \times q}$. However, the PML algorithm does not have direct access to the complete label matrix during training. Inspired by the low-rank assumption commonly used in multi-label learning, where labels exhibit inherent correlations, we similarly assume that the ground-truth label matrix $\tilde{Y}$ is low-rank. Under this assumption, the complex label matrix can be effectively represented in a lower-dimensional label subspace. This subspace can be approximated by the product of two smaller matrices, thereby simplifying the original matrix and reducing computational complexity:

$$\tilde{Y} \simeq VP \quad (3.1)$$

Here, $V \in \mathbb{R}^{n \times k}$ represents the sample representations in the k -dimensional latent label subspace, and $P \in \mathbb{R}^{k \times q}$ encodes the correlations between the original labels and the latent labels. Each original label is influenced by all k latent labels, implying a higher-order correlation structure among the original labels. This formulation captures complex label dependencies through the interaction between the latent space and observed labels.

To effectively perform matrix decomposition, we minimize the reconstruction error between the candidate label matrix Y and the product of V and P , as shown below:

$$\min_{V, P} \|Y - VP\|_F^2 + R(V, P) \quad (3.2)$$

where $\|\cdot\|_F$ denotes the Frobenius norm, $R(V, P)$ denotes the regularization for controlling model complexity.

In existing PML methods, most efforts have focused on addressing noise in the label space, while the impact of noise in the feature space is often overlooked. However, in real-world applications, feature information is frequently corrupted by outliers and noise similar to labels. To address this issue, we introduce a second subspace representation in the prediction model: the latent feature subspace.

By learning the feature correlation matrix $U \in \mathbb{R}^{d \times m}$, the original feature space is mapped to a low-dimensional feature subspace, thereby providing a compact and discriminative feature information and reducing the negative impact of noise feature information. Here m is the dimension of the feature subspace and the latent feature representation in the m -dimensional subspace can be expressed as XU .

We further introduce a model coefficient matrix $W \in \mathbb{R}^{m \times k}$, which can map samples from latent feature subspace to latent label subspaces, as follows:

$$\min_{V, P, W} \|Y - VP\|_F^2 + \|X - XUW\|_F^2 + R(V, P) \quad (3.3)$$

Note that the observed label matrix Y generally contains irrelevant or noisy labels. To better capture the underlying structure, we decompose Y into two components: a sparse noise label matrix Q and a label confidence matrix $\bar{Y}$. In many real-world scenarios, noise labels are

assumed to be sparsely distributed, which motivates the use of the ℓ_1 -norm to constrain Q . Meanwhile, due to label correlations, the label confidence matrix $\bar{Y}$ is assumed to be low-rank. To enforce this property, we apply a nuclear norm regularization to $\bar{Y}$. Thus, the observed label matrix can be modeled as the sum of a sparse noise matrix and a low-rank ground-truth matrix, expressed as:

$$\begin{aligned} \min_{P, Q} \quad & \|\bar{Y}\|_* + \lambda \|Q\|_1 \\ \text{s.t.} \quad & Y = \bar{Y} + Q \end{aligned} \quad (3.4)$$

Here, the parameter $\lambda \in [0, 1]$ controls the balance between the sparse noise label matrix Q and the low-rank ground-truth label matrix $\bar{Y}$. By incorporating the decomposition from Eq. (3.4) into the original matrix reconstruction objective in Eq. (3.3), we derive the following unified optimization formulation:

$$\begin{aligned} \min_{W, V, P, Q, E} \quad & \|\bar{Y} - VP\|_F^2 + \lambda_1 \|V - XUW\|_F^2 + \lambda_2 \|E\|_* + \lambda_3 \|Q\|_1 + \|W\|_F^2 \\ \text{s.t.} \quad & Y = \bar{Y} + Q \end{aligned} \quad (3.5)$$

To better capture the complex structure of the label space, we consider that the ground-truth labels lie in a union of multiple low-dimensional subspaces. This perspective enables the model to preserve the diversity and granularity of the underlying data structure. Accordingly, the following optimization problem can be formulated to simultaneously recover the true labels and identify the subspace structures:

$$\begin{aligned} \min_{W, Z, Q, V, P} \quad & \|\bar{Y} - VP\|_F^2 + \lambda_1 \|V - XUW\|_F^2 + \lambda_2 \|P\|_* + \lambda_3 \|Q\|_1 + \lambda_4 \|W\|_F^2 \\ \text{s.t.} \quad & Y = \bar{Y} + Q, \quad Z \geq 0 \end{aligned} \quad (3.6)$$

In multi-label data, each sample is associated with multiple labels simultaneously, and these labels are inherently determined by the sample's features. Consequently, the local topological structure in the feature space should be consistent with that in the label space (Zhu, Kwok and Zhou, 2017). To preserve this structural alignment, we incorporate manifold regularization using the graph Laplacian operator. This regularization encourages similar samples in the feature space to remain close in the ground-truth label space, thereby enhancing the ability to distinguish true labels from noisy candidates. By leveraging this local consistency, the model is better guided in identifying the accurate label assignments from the noisy candidate label matrix.

Let $S \in \mathbb{R}^{n \times n}$ denote the pairwise similarity matrix, where each entry S_{ij} measures the similarity between instances x_i and x_j . If two instances are closely related, i.e., S_{ij} is large, they are more likely to be similar in the shared latent subspace. This similarity can be exploited to enhance the accuracy of feature selection and label recovery. To model this local structure, we construct a k -NN graph G , where each node represents an instance and edges connect it to its k most similar neighbors based on feature similarity. The similarity matrix S is then defined as:

$$S_{ij} = \begin{cases} \exp\left(-\frac{\|x_i - x_j\|_2^2}{\sigma^2}\right), & \text{if } x_j \in \mathcal{N}_p(x_i) \\ 0, & \text{otherwise} \end{cases} \quad (3.7)$$

Here, $\mathcal{N}_p(x_i)$ denotes the set of k -nearest neighbors of instance x_i . Incomplete or noisy label spaces can disrupt the underlying semantic structure, making it difficult to accurately capture and represent label information. To address this, we introduce a smoothness assumption into the framework: if two samples are close in the feature space, they should also remain close in the shared latent subspace. This assumption helps preserve local consistency and guides the model to recover coherent label representations, even in the presence of missing or corrupted labels. This leads to the incorporation of a manifold regularization term based on the graph Laplacian, formulated as:

$$\sum_{i=1}^n \sum_{j=1}^n S_{ij} \left\| \frac{v_{ij}}{\sqrt{E_{ii}}} - \frac{v_j}{\sqrt{E_{jj}}} \right\|_2^2 = \text{Tr}(V^T L V) \quad (3.8)$$

where $L = E^{-\frac{1}{2}}(E - S)E^{-\frac{1}{2}}$ is the graph Laplacian matrix, and E is the diagonal matrix with $E_{ii} = \sum_{j=1}^n S_{ij}$. By combining the manifold regularization term in Eq.(3.7) and the subspace learning formulation in Eq.(3.8) into the unified objective in Eq.(3.6), we arrive at the final optimization problem:

$$\begin{aligned} \min_{W, Q, V, P, \bar{Y}} \theta(W, Q, V, P, \bar{Y}) &= \frac{1}{2} \{ \|\bar{Y} - VP\|_F^2 + \lambda_1 \text{Tr}(V^T L V) + \lambda_3 \|Q\|_1 + \lambda_4 (\|W\|_F^2 + \|V\|_F^2 + \|P\|_F^2) \} \\ \text{s.t. } Y &= \bar{Y} + Q, U^T U = I \end{aligned} \quad (3.9)$$

4 Optimization solution

To solve the optimization problem in Equation (3.9), we adopt the block coordinate descent (BCD) method. This approach iteratively updates each of the variables, while holding the others fixed at each step (Ruzgas, Arnastauskaite and Juskaite, 2021) (Li, Tang, Zhang and Yang, 2020) (Eligo, 2024). By decomposing the problem into smaller subproblems, each with respect to a single block of variables, the BCD method efficiently finds a local minimum of the overall objective.

Update P: Equation (3.9) can be simplified with respect to P , leading to the following subproblem:

$$\min_P \frac{1}{2} \|\bar{Y} - VP\|_F^2 + \frac{1}{2} \lambda_4 \|P\|_F^2 \quad (4.1)$$

The solution is obtained as

$$P = (V^T V + \lambda_4 I)^{-1} V^T \bar{Y} \quad (4.2)$$

Update V: The matrix V is computed by minimizing the following objective function, derived from Equation (3.9) with other variables fixed:

$$\min_V \|\bar{Y} - VP\|_F^2 + \lambda_1 \|V - XUW\|_F^2 + \lambda_2 \text{Tr}(V^T L V) + \lambda_4 \|V\|_F^2 \quad (4.3)$$

The objective function (4.3) is differentiable, allowing V to be optimized using the standard gradient descent algorithm.

$$\begin{aligned} \frac{\partial \theta}{\partial V} &= (\lambda_1 + \lambda_4) V + \lambda_2 L V + V P P^T - Y P^T - \lambda_1 X U W \\ V &:= V - \lambda_V \frac{\partial \theta}{\partial V} \end{aligned} \quad (4.4)$$

where λ_V is the step size used in the gradient descent update. To ensure convergence and stable progress, λ_V is determined using the Armijo rule (Lyu et al., 2020).

Update U: To update U , we fix the other variables and minimize the following objective function:

$$\begin{aligned} \min_U \quad & \frac{1}{2} \left(\lambda_1 \|V - XUW\|_F^2 \right) \\ \text{s.t.} \quad & U^T U = I \end{aligned} \quad (4.5)$$

We can compute that $\frac{\partial \theta}{\partial U} = (X^T XUWW^T - X^T VW^T)$. The update rule is obtained as:

$$U := U - \lambda_U \frac{\partial \theta}{\partial U} \quad (4.6)$$

To satisfy the constraint $U^T U = I$, we project each row of U onto the unit norm ball after each iteration:

$$U_{i,:} \leftarrow \frac{U_{i,:}}{\|U_{i,:}\|} \quad (4.7)$$

Update W: To update the feature projection matrix W , we minimize the part of the objective function related to the manifold regularization and feature subspace learning. The corresponding optimization problem is:

$$\min_W \quad \frac{1}{2} (\|V - XUW\|_F^2 + \lambda_4 \|W\|_F^2) \quad (4.8)$$

We have that:

$$W = \lambda_2 (\lambda_2 U^T X^T XU + \lambda_4 I)^{-1} U^T X^T V \quad (4.9)$$

Update $\bar{Y}$: The updated formula for variable $\bar{Y}$ is as follows

$$\min_{\bar{Y}} \quad \frac{1}{2} \left\| \bar{Y} - VP \right\|_F^2 + \mu_1 \left\| Y - \bar{Y} - Q + \frac{Y_1}{\mu_1} \right\|_F^2 \quad (4.10)$$

The solution of $\bar{Y}$ is obtained as:

$$\begin{aligned} \frac{\partial \theta}{\partial \bar{Y}} &= (1 + \mu_1 \bar{Y}) - VP - \mu_1 Y + \mu_1 Q - Y_1 \\ \bar{Y} &:= \bar{Y} - \lambda_{\bar{Y}} \frac{\partial \theta}{\partial \bar{Y}} \end{aligned} \quad (4.11)$$

Update Q: The updated formula of variable Q is as follows:

$$\min_Q \quad \frac{1}{2} \left\| Q - \left(Y - \bar{Y} + \frac{Y_1}{\mu_1} \right) \right\|_F^2 + \frac{\lambda_3}{\mu_1} \|Q\|_1 \quad (4.12)$$

This objective consists of a Frobenius norm term and an ℓ_1 -norm term, promoting sparsity in Q . The optimal solution can be obtained using element-wise soft-thresholding (Cai, Candes and Shen, 2010), which is commonly used for problems of this form:

$$Q_{k+1} = \arg \min_Q \varepsilon \|Q\|_1 + \frac{1}{2} \|Q - G_k\|_F^2 = \delta_\varepsilon[G_k] \quad (4.13)$$

5 Experiment settings

5.1 Experiment settings

In this paper, we conducted experiments on nine real-world multi-label datasets. For each multi-label dataset, the label noise rate was adjusted within the range of 0% to 80%, following the noise settings previously defined, the parameters $\lambda_1, \lambda_2, \lambda_3$, and λ_4 range from $[2^{-6}, 2^{-4}, \dots, 2^6]$. Each algorithm was conducted the five-fold cross-validation (Tarekegn, Michalak and Giacobini, 2020)(Suman and Arindam, 2025). The detailed information of the datasets used in the experiments is presented in Table 1. In the table, N represents the number of samples, D represents the number of features, L represents the number of labels, "*Card*" indicated the average cardinality of labels associated with the sample, and "*Domain*" represents the category of the dataset. Additionally, the classification performance of the PMLFS algorithm and the comparison algorithms was evaluated across four dimensions: AP (Average Precision), HL (Hamming Loss), OE (One Error), and RL (Ranking Loss). Among these, lower values of HL, OE, and RL indicate better model performance, while a higher value of AP indicates better model performance.

To assess the performance of the PMLSR method, we compare it with six widely recognized multi-label methods. These algorithms include: MLMLFS (Zhu, Xu, Hu, Zhang and Zhao, 2018), PARM (Wang and Zhang, 2020), PARTICAL-MA (Xie and Huang, 2018), PARTICAL-VLS (Xie and Huang, 2018), CSFS (Chang, Nie, Yang and Huang, 2014) and SCLS (Lee and Kim, 2017), the parameters are set according to the recommendations in the respective literature. Specific datasets information can be found in Table 1.

Table 1: Description of the experimental datasets

Datasets	N	D	L	<i>Domain</i>
birds	645	258	19	audio
business	5000	438	30	text
computer	5000	681	33	text
emotions	593	72	6	music
education	5000	550	33	text
entertainment	5000	640	21	text
health	5000	612	32	text
image	2000	294	5	image
yeast	2417	103	14	biology

5.2 Experiment result

Tables 2-5 present the experimental results of seven algorithms on nine PML datasets. The upward arrow symbol($\uparrow$) in the table header indicates that a higher value of the metric corresponds to better classification performance, while the downward arrow symbol($\downarrow$) indicates that a lower value corresponds to better performance. The symbol ϵ represents the label noise rate. The metrics are selected for label noise rates of 30%, 50%, and 70%. Among them, bold black fonts indicate the best algorithm results. Some conclusions can be drawn from the tables.

Table 2: Experimental results on the evaluation metric AP($\uparrow$)

Datasets	ε	PMLSR	MLMFS	PARM	PAR_map	PAR_vls	CSFS	SCLS
birds	30%	0.7041(1)	0.6144(5)	0.6058(7)	0.6809(2)	0.6545(3)	0.6181(4)	0.6088(6)
	50%	0.7053(1)	0.6429(4)	0.6426(5)	0.6967(2)	0.6846(3)	0.6180(6)	0.6143(7)
	70%	0.7203(1)	0.6006(5)	0.6271(4)	0.7133(2)	0.7078(3)	0.5956(6)	0.5944(7)
business	30%	0.8689(1)	0.8660(3)	0.8582(6)	0.8653(4)	0.8448(7)	0.8677(2)	0.8598(5)
	50%	0.8657(2)	0.8494(7)	0.8566(6)	0.8784(1)	0.8610(4)	0.8642(3)	0.8592(5)
	70%	0.8646(1)	0.8609(3)	0.8571(6)	0.8599(4)	0.8540(7)	0.8617(2)	0.8583(5)
computer	30%	0.6340(1)	0.5802(6)	0.5993(3)	0.6248(2)	0.4454(7)	0.5831(5)	0.5849(4)
	50%	0.6441(1)	0.6143(3)	0.5970(4)	0.6391(2)	0.5039(7)	0.5774(6)	0.5840(5)
	70%	0.6288(2)	0.5790(5)	0.5930(3)	0.6530(1)	0.5592(7)	0.5757(6)	0.5804(4)
emotions	30%	0.7370(1)	0.6551(5)	0.7058(4)	0.7285(2)	0.7127(3)	0.6405(6)	0.6231(7)
	50%	0.7564(1)	0.7480(2)	0.7035(5)	0.7468(3)	0.7344(4)	0.6141(7)	0.6239(6)
	70%	0.7494(1)	0.6142(6)	0.6476(3)	0.6775(2)	0.6250(4)	0.6179(5)	0.5859(7)
education	30%	0.5572(1)	0.4956(6)	0.4995(4)	0.5560(2)	0.5257(3)	0.4946(7)	0.4993(5)
	50%	0.6478(1)	0.4857(7)	0.5015(4)	0.6317(2)	0.5649(3)	0.4916(5)	0.4909(6)
	70%	0.6050(1)	0.4818(6)	0.4927(4)	0.5941(2)	0.5819(3)	0.4857(7)	0.4919(5)
entertainment	30%	0.5532(2)	0.5302(5)	0.4892(7)	0.5781(1)	0.5224(6)	0.5382(4)	0.5466(3)
	50%	0.5533(1)	0.5459(3)	0.4929(7)	0.5506(2)	0.5327(5)	0.5386(4)	0.5457(6)
	70%	0.5387(1)	0.5289(4)	0.4896(7)	0.5089(5)	0.5045(6)	0.5305(3)	0.5374(2)
health	30%	0.6656(1)	0.6377(5)	0.6252(6)	0.6449(3)	0.4606(7)	0.6602(2)	0.6424(4)
	50%	0.6580(3)	0.6621(2)	0.6228(6)	0.7380(1)	0.5686(7)	0.6538(4)	0.6462(5)
	70%	0.6550(1)	0.6348(5)	0.6213(6)	0.6412(4)	0.4915(7)	0.6425(2)	0.6418(3)
image	30%	0.7318(1)	0.6762(5)	0.5426(7)	0.7165(3)	0.7205(2)	0.6832(4)	0.6735(6)
	50%	0.7278(1)	0.7113(2)	0.6343(7)	0.6825(5)	0.7022(3)	0.6796(4)	0.6697(6)
	70%	0.6753(1)	0.6513(3)	0.5251(7)	0.6101(5)	0.6049(6)	0.6632(2)	0.6450(4)
yeast	30%	0.7220(1)	0.6519(7)	0.7163(2)	0.7129(3)	0.7065(4)	0.6596(5)	0.6565(6)
	50%	0.7461(1)	0.7398(3)	0.7139(5)	0.7455(2)	0.7377(4)	0.6563(6)	0.6527(7)
	70%	0.7082(1)	0.6476(6)	0.6763(4)	0.7053(2)	0.6928(3)	0.6505(5)	0.6471(7)

Table 3: Experimental results on the evaluation metric OE($\downarrow$)

Datasets	ε	PMLSR	MLMFS	PARM	PAR_map	PAR_vls	CSFS	SCLS
birds	30%	0.3530(1)	0.4458(6)	0.4768(7)	0.3868(3)	0.3791(2)	0.4356(5)	0.4248(4)
	50%	0.3492(1)	0.4597(6)	0.4737(7)	0.3643(2)	0.3798(3)	0.4285(5)	0.4280(4)
	70%	0.3023(1)	0.4404(5)	0.4613(7)	0.3766(3)	0.3638(2)	0.4486(6)	0.4322(4)
business	30%	0.1317(1)	0.1393(5)	0.1367(4)	0.1460(7)	0.1340(2)	0.1401(6)	0.1362(3)
	50%	0.1180(1)	0.1352(5)	0.1367(7)	0.1230(3)	0.1214(2)	0.1318(4)	0.1365(6)
	70%	0.1050(1)	0.1328(5)	0.1367(7)	0.1070(3)	0.1046(2)	0.1321(4)	0.1357(6)
computer	30%	0.4320(2)	0.4480(6)	0.4820(7)	0.4340(3)	0.3820(1)	0.4356(4)	0.4358(5)
	50%	0.4290(2)	0.4420(5)	0.4803(7)	0.4310(3)	0.4120(1)	0.4430(6)	0.4348(4)
	70%	0.4367(2)	0.4408(4)	0.4820(7)	0.4430(6)	0.3740(1)	0.4415(5)	0.4381(3)
emotions	30%	0.3317(2)	0.3366(3)	0.4158(7)	0.3577(4)	0.3190(1)	0.3683(5)	0.3911(6)
	50%	0.3436(1)	0.3525(3)	0.4208(7)	0.3445(2)	0.3593(4)	0.3896(6)	0.3733(5)
	70%	0.3433(1)	0.3827(3)	0.4257(5)	0.3773(2)	0.4454(7)	0.3869(4)	0.4386(6)
education	30%	0.4917(2)	0.6633(7)	0.6550(4)	0.4940(3)	0.4680(1)	0.6592(5)	0.6622(6)
	50%	0.4877(2)	0.6468(4)	0.6540(5)	0.5020(3)	0.4760(1)	0.6639(6)	0.6677(7)

	70%	0.5290(1)	0.6763(7)	0.6650(4)	0.6090(3)	0.5840(2)	0.6688(6)	0.6655(5)
entertainment	30%	0.6117(2)	0.6376(5)	0.6937(7)	0.6070(1)	0.6430(6)	0.6285(4)	0.6173(3)
	50%	0.6097(1)	0.6411(5)	0.6877(7)	0.6180(2)	0.6530(6)	0.6264(4)	0.6201(3)
	70%	0.6203(2)	0.6398(5)	0.6900(7)	0.5800(1)	0.6460(6)	0.6380(4)	0.6295(3)
health	30%	0.4220(2)	0.4643(5)	0.4673(6)	0.3200(1)	0.5950(7)	0.4276(3)	0.4593(4)
	50%	0.4317(2)	0.4660(5)	0.4723(6)	0.3230(1)	0.5870(7)	0.4388(3)	0.4493(4)
	70%	0.4341(1)	0.4684(5)	0.4717(6)	0.4490(2)	0.6610(7)	0.4527(4)	0.4515(3)
image	30%	0.4040(2)	0.4176(4)	0.6820(7)	0.5750(6)	0.3975(1)	0.4075(3)	0.4208(5)
	50%	0.4020(1)	0.4377(6)	0.4760(7)	0.4350(5)	0.4025(2)	0.4134(3)	0.4282(4)
	70%	0.4034(2)	0.4501(5)	0.7100(7)	0.4086(3)	0.3800(1)	0.4382(4)	0.4594(6)
yeast	30%	0.2350(2)	0.2556(6)	0.2530(5)	0.2588(7)	0.2402(3)	0.2274(1)	0.2479(4)
	50%	0.2148(1)	0.2432(6)	0.2541(7)	0.2422(5)	0.2381(4)	0.2266(2)	0.2274(3)
	70%	0.2231(1)	0.2321(5)	0.2672(7)	0.2340(6)	0.2312(4)	0.2278(3)	0.2272(2)

Table 4: Experimental results on the evaluation metric HL($\downarrow$)

Datasets	ε	PMLSR	MLMFS	PARM	PAR_map	PAR_vls	CSFS	SCLS
birds	30%	0.0593(1)	0.0601(2)	0.0680(6)	0.0705(7)	0.0609(4)	0.0606(3)	0.0610(5)
	50%	0.0566(2)	0.0605(5)	0.0683(7)	0.0589(3)	0.0512(1)	0.0608(6)	0.0596(4)
	70%	0.0492(1)	0.0585(2)	0.0689(6)	0.0713(7)	0.0667(5)	0.0630(4)	0.0603(3)
business	30%	0.0185(1)	0.0283(4)	0.0291(6)	0.0277(2)	0.0508(7)	0.0282(3)	0.0287(5)
	50%	0.0280(2)	0.0295(6)	0.0288(5)	0.0260(1)	0.0511(7)	0.0284(3)	0.0287(4)
	70%	0.0276(1)	0.0300(4)	0.0293(3)	0.0286(2)	0.0535(7)	0.0384(5)	0.0386(6)
computer	30%	0.0353(1)	0.0403(4)	0.0432(5)	0.0435(6)	0.0452(7)	0.0391(2)	0.0395(3)
	50%	0.0393(2)	0.0430(5)	0.0435(6)	0.0404(4)	0.0449(7)	0.0385(1)	0.0396(3)
	70%	0.0392(1)	0.0397(2)	0.0441(4)	0.0425(3)	0.0451(5)	0.0496(7)	0.0459(6)
emotions	30%	0.2255(1)	0.2386(4)	0.2657(7)	0.2353(3)	0.2307(2)	0.2464(5)	0.2599(6)
	50%	0.2379(1)	0.2595(3)	0.2880(7)	0.2486(2)	0.2680(4)	0.2757(6)	0.2724(5)
	70%	0.2407(1)	0.3809(5)	0.4629(7)	0.2479(2)	0.2871(3)	0.3651(4)	0.4133(6)
education	30%	0.0423(1)	0.0908(5)	0.0980(6)	0.0452(3)	0.0430(2)	0.0876(4)	0.0993(7)
	50%	0.0378(1)	0.0795(5)	0.0687(4)	0.0435(3)	0.0426(2)	0.0912(6)	0.0974(7)
	70%	0.0432(2)	0.1000(4)	0.1908(5)	0.0482(3)	0.0399(1)	0.2136(6)	0.1853(7)
entertainment	30%	0.0622(1)	0.0642(4)	0.0675(5)	0.1204(7)	0.0835(6)	0.0633(2)	0.0641(3)
	50%	0.0636(2)	0.0644(4)	0.0675(5)	0.1159(7)	0.0819(6)	0.0629(1)	0.0639(3)
	70%	0.0326(1)	0.0645(3)	0.0682(5)	0.1154(7)	0.0811(6)	0.0639(2)	0.0646(4)
health	30%	0.0416(1)	0.0482(5)	0.0497(6)	0.0473(3)	0.0653(7)	0.0455(2)	0.0476(4)
	50%	0.0397(1)	0.0485(5)	0.0496(6)	0.0467(2)	0.0615(7)	0.0477(4)	0.0475(3)
	70%	0.0401(1)	0.0485(4)	0.0496(5)	0.0464(2)	0.0649(7)	0.0477(3)	0.0517(6)
image	30%	0.2086(1)	0.2184(4)	0.2600(6)	0.4180(7)	0.2195(5)	0.2155(3)	0.2105(2)
	50%	0.1890(1)	0.2505(6)	0.2824(7)	0.1930(2)	0.2035(3)	0.2455(5)	0.2375(4)
	70%	0.1867(1)	0.3034(5)	0.3642(7)	0.2385(3)	0.2230(2)	0.2967(4)	0.3148(6)
yeast	30%	0.1907(1)	0.1938(2)	0.2315(7)	0.2159(6)	0.2098(5)	0.1943(3)	0.1997(4)
	50%	0.2004(1)	0.2078(5)	0.2488(7)	0.2056(3)	0.2039(2)	0.2063(4)	0.2136(6)
	70%	0.1950(1)	0.3020(6)	0.3573(7)	0.2124(3)	0.2100(2)	0.2915(4)	0.3009(5)

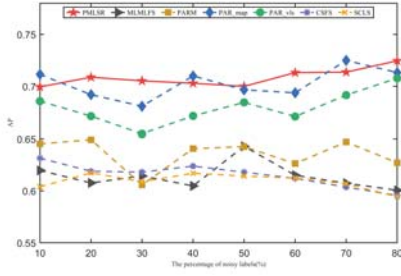
Table 5: Experimental results on the evaluation metric RL($\downarrow$)

Datasets	ε	PMLSR	MLMFS	PARM	PAR_map	PAR_vls	CSFS	SCLS
----------	---------------	-------	-------	------	---------	---------	------	------

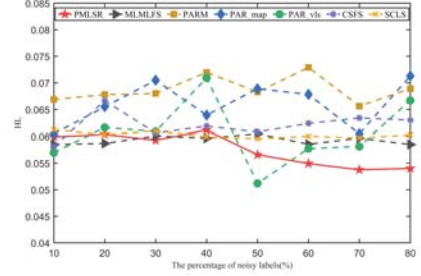
birds	30%	0.1645(2)	0.1749(6)	0.1691(4)	0.1098(1)	0.1706(5)	0.1835(7)	0.1660(3)
	50%	0.1542(2)	0.1550(3)	0.1662(6)	0.1200(1)	0.1868(7)	0.1608(4)	0.1613(5)
	70%	0.1502(1)	0.1653(5)	0.1571(4)	0.1518(2)	0.1728(7)	0.1523(3)	0.1704(6)
business	30%	0.0465(1)	0.0531(6)	0.0520(5)	0.0513(4)	0.0930(7)	0.0484(2)	0.0512(3)
	50%	0.0428(1)	0.0565(6)	0.0531(4)	0.0472(2)	0.0860(7)	0.0518(3)	0.0535(5)
	70%	0.0408(1)	0.0561(4)	0.0586(6)	0.0532(2)	0.0894(7)	0.0544(3)	0.0570(5)
computer	30%	0.0968(1)	0.1066(6)	0.1062(4)	0.0973(2)	0.1333(7)	0.0987(3)	0.1064(5)
	50%	0.0840(1)	0.1042(5)	0.1069(6)	0.0918(2)	0.1483(7)	0.1013(4)	0.1002(3)
	70%	0.0826(1)	0.1037(4)	0.1120(6)	0.0956(2)	0.1418(7)	0.1053(5)	0.1015(3)
emotions	30%	0.2404(2)	0.2519(4)	0.2592(7)	0.1808(1)	0.2532(6)	0.2432(3)	0.2525(5)
	50%	0.2001(1)	0.2207(2)	0.2759(6)	0.2248(3)	0.2323(4)	0.2766(7)	0.2614(5)
	70%	0.1607(1)	0.2735(4)	0.3128(7)	0.1782(2)	0.3035(6)	0.2677(3)	0.3000(5)
education	30%	0.0926(2)	0.1063(4)	0.1079(6)	0.0683(1)	0.0949(3)	0.1093(7)	0.1071(5)
	50%	0.0923(2)	0.1122(7)	0.1067(4)	0.0797(1)	0.0948(3)	0.1102(6)	0.1098(5)
	70%	0.0902(1)	0.1147(6)	0.1119(5)	0.0677(2)	0.0857(3)	0.1152(7)	0.1105(4)
entertainment	30%	0.1284(1)	0.1298(3)	0.1503(7)	0.1296(2)	0.1481(6)	0.1334(5)	0.1318(4)
	50%	0.1288(1)	0.1336(3)	0.1474(6)	0.1359(5)	0.1506(7)	0.1342(4)	0.1332(2)
	70%	0.1172(2)	0.1377(4)	0.1537(7)	0.1147(1)	0.1414(6)	0.1393(5)	0.1371(3)
health	30%	0.0589(1)	0.0688(3)	0.0815(6)	0.0653(2)	0.1086(7)	0.0725(4)	0.0754(5)
	50%	0.0564(1)	0.0697(3)	0.0827(6)	0.0672(2)	0.0912(7)	0.0767(5)	0.0747(4)
	70%	0.0517(1)	0.0797(3)	0.0865(6)	0.0703(2)	0.1038(7)	0.0812(5)	0.0810(4)
image	30%	0.2255(1)	0.2341(3)	0.4373(7)	0.3825(6)	0.2754(5)	0.2334(2)	0.2455(4)
	50%	0.2243(1)	0.2497(3)	0.4412(7)	0.3292(6)	0.2550(5)	0.2390(2)	0.2507(4)
	70%	0.2116(1)	0.2731(4)	0.4600(7)	0.3106(6)	0.2650(3)	0.2552(2)	0.2760(5)
yeast	30%	0.1919(1)	0.1973(5)	0.2046(7)	0.1722(2)	0.1963(4)	0.1927(3)	0.2038(6)
	50%	0.1601(1)	0.1888(6)	0.2077(7)	0.1776(3)	0.1882(5)	0.1759(2)	0.1782(4)
	70%	0.1154(1)	0.1845(5)	0.2125(7)	0.1717(2)	0.2050(6)	0.1821(3)	0.1842(4)

Firstly, on the nine datasets, the proportion of times that the PMLSR algorithm ranked first across all evaluation metrics was 77.8%. For each comparison algorithm, PMLSR significantly outperformed the PML method in 100%, 83.33%, and 91.67% of the cases, and significantly outperformed the MLL method in 96.3%, 83.33%, and 80% of the cases. For each evaluation metric, under the AP metric, the proportion of times that PMLSR ranked first was 85.19%; under the OE metric, it was 59.26%; under the HL metric, it was 85.19%; and under the RL metric, it was 81.48%.

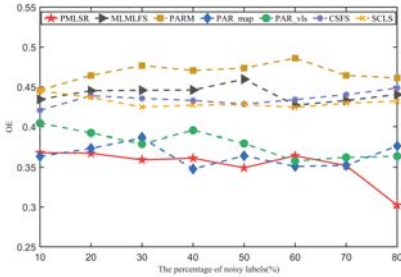
Secondly, Figures 1 and 2 illustrate the performance of different algorithms on the dataset birds and emotions under four evaluation metrics. The compared algorithms fluctuate with the increase in label noise rate. However, the PMLSR method demonstrates better stability in these fluctuations, and its performance does not significantly decline as the noise ratio rises, maintaining a relatively high level overall. The PMLSR method significantly outperforms PARM, SCLS, and CSFS across all evaluation metrics, especially as the label noise ratio increases. Not only does the PMLSR method continue to lead in metrics such as AP, HL, OE, and RL, but it also effectively reduces the amplitude of performance fluctuations caused by the introduction of noise. The experimental results demonstrate the robustness and superiority of the PMLSR



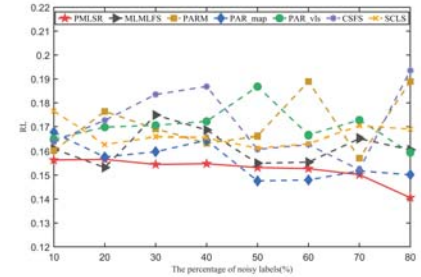
(a) AP



(b) HL

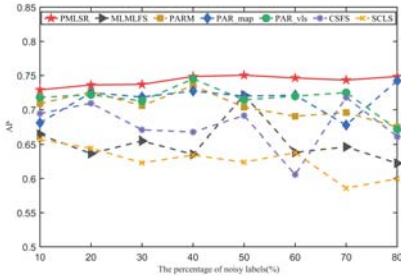


(c) OE

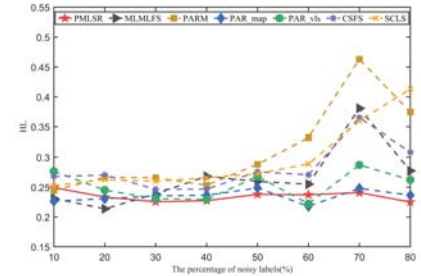


(d) RL

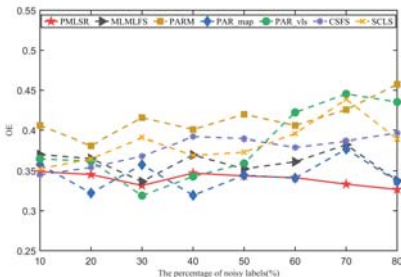
Figure 1: Performance comparison on the birds dataset



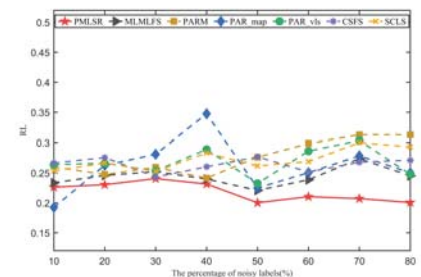
(a) AP



(b) HL



(c) OE



(d) RL

Figure 2: Performance comparison on the emotions dataset

method in dealing with label noise. It performs well across different evaluation metrics and maintains high classification accuracy even when faced with varying proportions of noisy label information. This further validates the practicality and reliability of the PMLSR method in the field of multi-label learning.

Finally, in order to analyze the significant differences in performance between the proposed method and other comparison algorithms, the Friedman test is conduct (Demsar, 2006). Table 4, 5, and 6 shows the Friedman statistic F_F and the critical values corresponding to each evaluation metric. The results indicate that at the significance level $\alpha = 0.05$, we reject the null hypothesis that all comparison algorithms have the same performance. This means that there are significant performance differences among the compared methods. Moreover, in order to further verify whether the PMLSR method outperforms other comparison methods in terms of performance, the Nemenyi test was employed. By comparing the differences between different algorithms, the test assesses whether there are significant differences in the performance of each method.

Table 6: Friedman statistic F_F ($N = 9, K = 7$) and critical values under different evaluation indicators ($\alpha = 0.05$)

ε	Evaluation metric	F_F	Critical value
30%	AP	28.9524	2.6864
	HL	23.3810	
	OE	23.9524	
	RL	30.5238	
50%	AP	33.1905	2.6864
	HL	22.9524	
	OE	32.3810	
	RL	27.8095	
70%	AP	24.2381	2.6864
	HL	22.6190	
	OE	26.8571	
	RL	33.8571	

In the tests, PMLSR is regarded as a control method to verify whether it can achieve or exceed the performance levels of other competing methods, thereby providing sufficient statistical support for the effectiveness and superiority of the algorithm. In the Nemenyi test, when the significance level is $\alpha = 0.05$, $q_\alpha = 2.78$, $CD = 2.6864(N = 9, K = 7)$, If the ranking difference between two algorithms is greater than or equal to one critical difference, that is, $CD = q_\alpha \sqrt{\frac{k(k+1)}{6N}}$, the algorithms are considered to have a significant difference. Figures 3-5 illustrate the CD diagrams for label noise rates of 30%, 50%, and 70%. If the two algorithms are connected by a red line, it indicates that the difference in average ranking is less than one CD , which is considered to be no significant difference. Although there is always no significant difference between the proposed PMLSR algorithm and the comparison algorithm, the PMLSR algorithm ranks first overall.

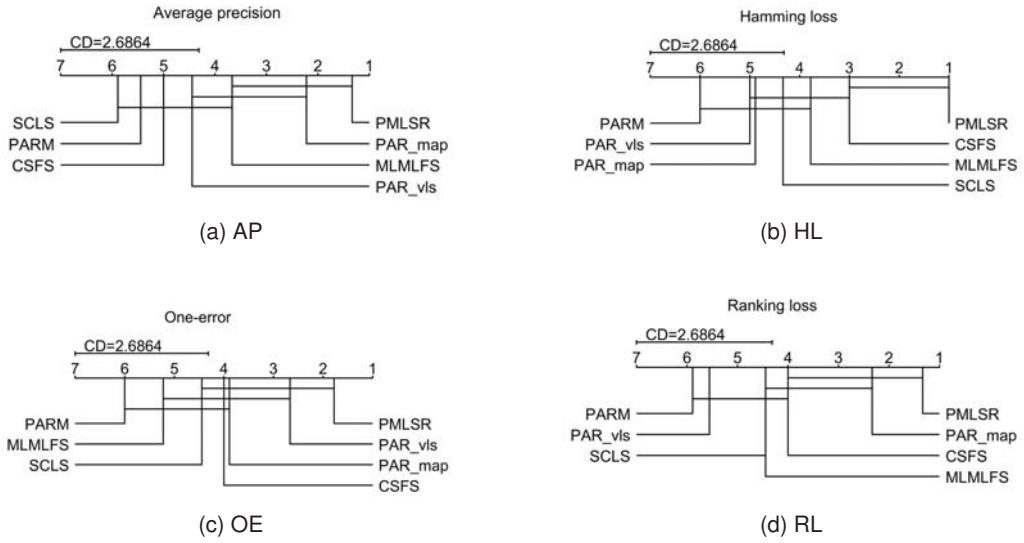


Figure 3: The label noise rate is 30%, and the results under the Bonferroni-Dunn test

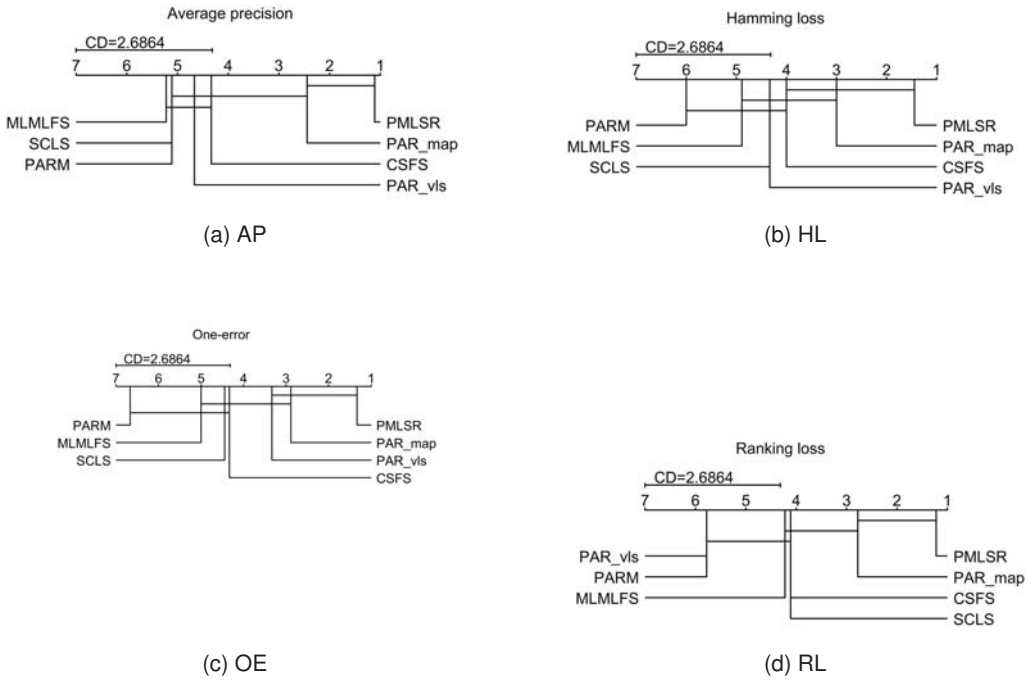


Figure 4: The label noise rate is 50%, and the results under the Bonferroni-Dunn test

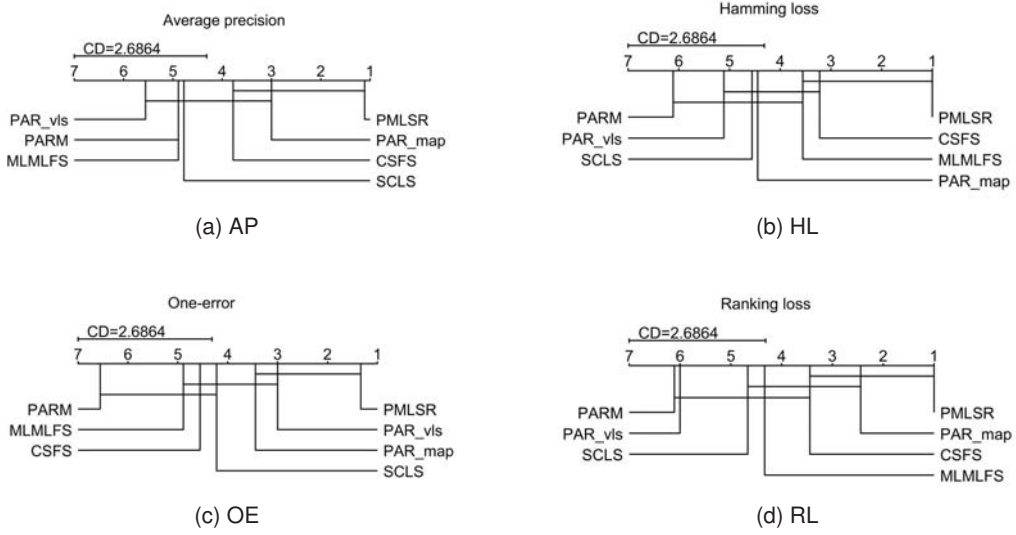
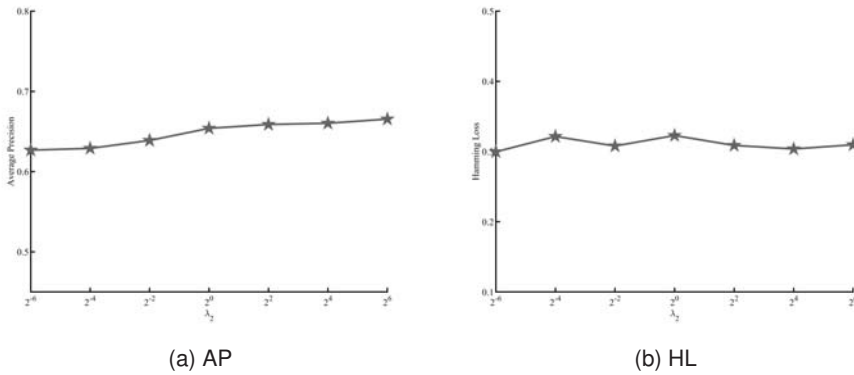
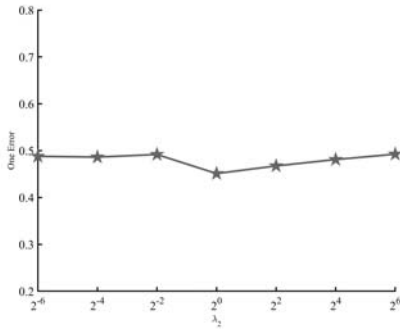


Figure 5: The label noise rate is 70%, and the results under the Bonferroni-Dunn test

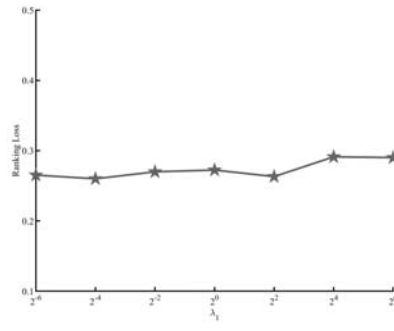
5.3 Parameter sensitivity

We analyze the four-trade off parameter in PMLSR, including λ_1 , λ_2 , λ_3 , and λ_4 . The influence of the four parameters on the experimental results is studied. Where λ_1 balances feature learning and matrix decomposition, λ_2 is the regularization parameter corresponding to manifold regularization, which controls the influence of instance correlation. λ_3 is the regularization parameter corresponding to the sparsity of noise labels, and λ_4 controls the degree differentiation between features. In order to analyze the parameter sensitivity of PMLSR, the image dataset and the label noise rate of 70% were selected. The four evaluation metrics were respectively AP, HL, OE, and RL. Only one parameter was changed, and the other parameters were fixed in the optimal setting $[2^{-6}, 2^{-4}, \dots, 2^6]$.



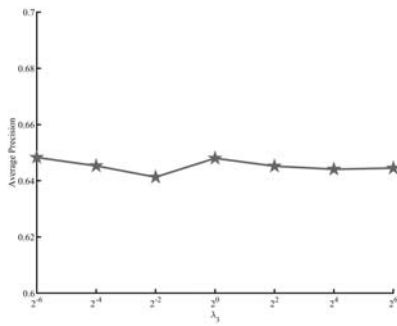


(a) OE

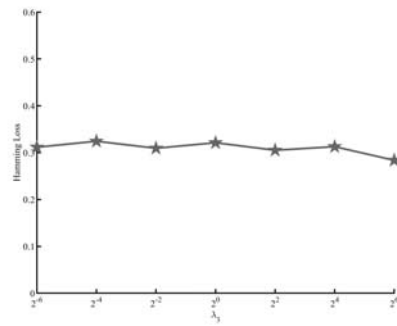


(b) RL

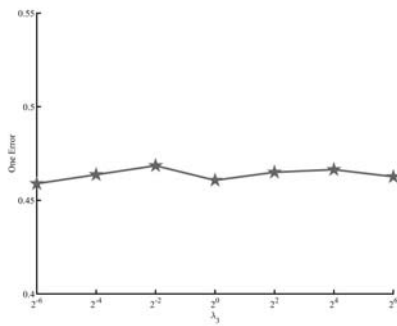
Figure 7: Sensitivity analysis of parameter λ_2 under four evaluation indicators on the dataset image, with a label noise rate of 70%



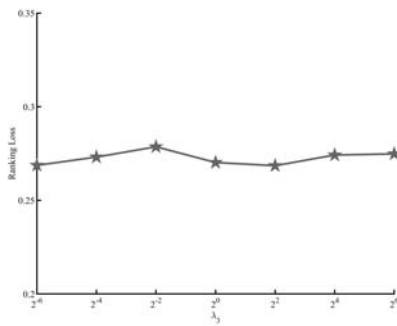
(a) AP



(b) HL



(c) OE



(d) RL

Figure 8: Sensitivity analysis of parameter λ_3 under four evaluation indicators on the dataset image, with a label noise rate of 70%

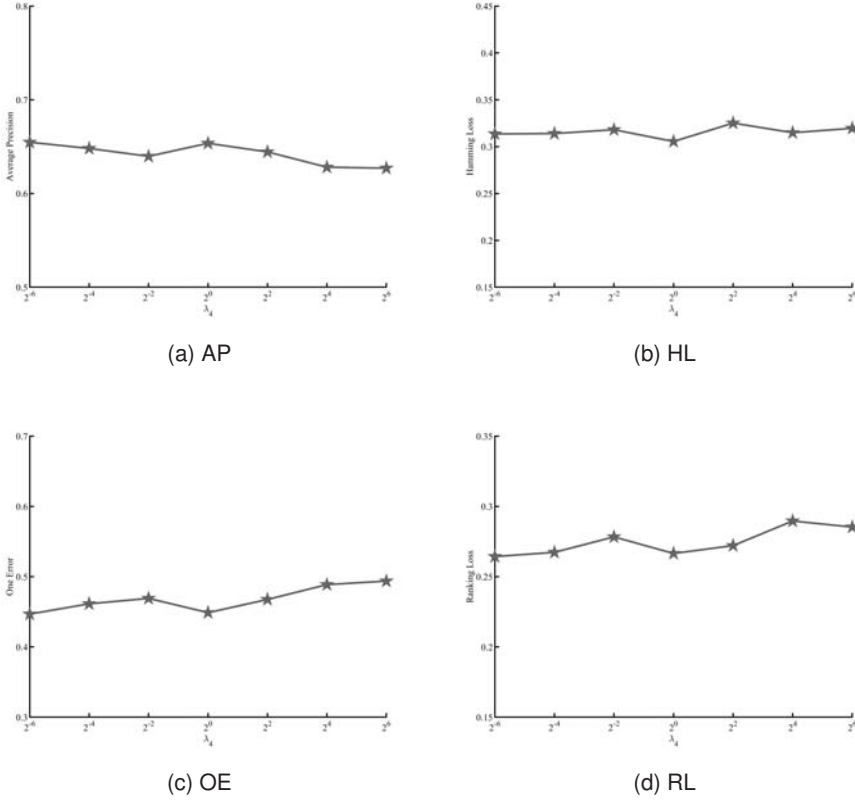


Figure 9: Sensitivity analysis of parameter λ_4 under four evaluation indicators on the dataset image, with a label noise rate of 70%

We can draw some conclusions from the Figures 7, 7, 8, and 9. When λ_1 varies from 2^{-6} to 2^0 , the performance of PMLSR remains stable. When λ_1 exceeds 2^0 , the performance of the model begins to decline with the increase of parameters, and the importance of irrelevant features becomes weaker and weaker. When λ_2 varies from 2^{-6} to 2^6 , the performance of PMLSR is stable and gradually improves. When λ_3 exceeds 2^2 , the performance degrades, because when λ_3 is large enough, the noise label matrix Q will becomes zero matrix. After λ_4 reaches its peak at 2^0 , the performance begins to deteriorate.

6 Conclusions

In this paper, we proposed Partial Multi-label Learning based on Shared Subspace Representation (PMLSR), a novel framework designed to address the challenges of partial multi-label learning. By explicitly modeling the correlation between features and ground-truth labels, PMLSR learns distinguishable shared subspaces from redundant labels, providing a more accurate representation of the underlying data structure. The framework incorporates robust mechanisms to handle noise in both the feature and label spaces, ensuring resilience to real-world data imperfections. Additionally, the introduction of the Laplacian and low-rank decom-

position techniques effectively mitigates the negative impact of redundant labels, focusing on the most informative aspects of the data. Through feature subspace mapping and orthogonal subspace projection, PMLSR ensures consistency between the label subspace and the feature space, enhancing the overall learning performance. Extensive empirical validation on multiple datasets demonstrates the effectiveness of PMLSR, showing significant improvements over existing methods.

In summary, PMLSR offers a comprehensive and robust solution to the challenges of partial multi-label learning, making it a valuable tool for real-world applications. Future work will explore extending this framework to other related learning paradigms and further optimizing its computational efficiency.

References

- Cai, J. F., Candes, E. J. and Shen, Z. W. 2010. A singular value thresholding algorithm for matrix completion, *SIAM Journal on optimization* **20**(4): 1956–1982.
- Chang, X. J., Nie, F. P., Yang, Y. and Huang, H. 2014. A convex formulation for semi-supervised multi-label feature selection, *Proceedings of the AAAI conference on artificial intelligence*, Vol. 28.
- Demsar, J. 2006. Statistical comparisons of classifiers over multiple data sets, *Journal of Machine learning research* **7**: 1–30.
- Eligo, W. M. 2024. High-dimensional data and the bias variance tradeoff in model selection, *International Journal of Applied Mathematics and Statistics* **63**: 17–33.
- Feng, L. and An, B. 2019. Partial label learning with self-guided retraining, *Proceedings of the AAAI conference on artificial intelligence*, Vol. 33, pp. 3542–3549.
- Feng, L., An, B. and He, S. 2019. Collaboration based multi-label learning, *Proceedings of the AAAI conference on artificial intelligence*, Vol. 33, pp. 3550–3557.
- Lee, J. S. and Kim, D. W. 2017. Scls: Multi-label feature selection based on scalable criterion for large label set, *Pattern Recognition* **66**: 342–352.
- Li, Z. C., Tang, J. H., Zhang, L. Y. and Yang, J. 2020. Weakly-supervised semantic guided hashing for social image retrieval, *International Journal of Computer Vision* **128**(8): 2265–2278.
- Lyu, G. Y., Feng, S. H., Li, Y. D., Jin, Y., Dai, G. J. and Lang, C. Y. 2020. Hera: partial label learning by combining heterogeneous loss with sparse and low-rank regularization, *ACM Transactions on Intelligent Systems and Technology (TIST)* **11**(3): 1–19.
- Ruzgas, T., Arnastauskaite, J. and Juskaite, I. 2021. Algorithm for discovery of sequential patterns in big data and application on price of brent crude oil, *International Journal of Applied Mathematics and Statistics* **60**(1): 24–30.

- Suman, K. and Arindam, G. 2025. Analyzing interval censored data using different parametric model, *International Journal of Mathematics Computation* **36**: 33–50.
- Sun, L. J., Feng, S. H., Lyu, G. Y., Zhang, H. and Dai, G. 2021. Partial multi-label learning with noisy side information, *Knowledge and Information Systems* **63**(2): 541–564.
- Tarekegn, A. N., Michalak, K. and Giacobini, M. 2020. Cross-validation approach to evaluate clustering algorithms: An experimental study using multi-label datasets, *SN Computer Science* **1**(5): 263.
- Wang, W. and Zhang, M. L. 2020. Semi-supervised partial label learning via confidence-rated margin maximization, *Advances in neural information processing systems* **33**: 6982–6993.
- Xie, M. K. and Huang, S. J. 2018. Partial multi-label learning, *Proceedings of the AAAI conference on artificial intelligence*, Vol. 32, pp. 4302–4309.
- Xie, M. K. and Huang, S. J. 2022. Partial multi-label learning with noisy label identification, *IEEE Transactions on Pattern Analysis and Machine Intelligence* **44**(7): 3676–3687.
- Xu, L. L., Wang, Z., Shen, Z. F., Wang, Y. B. and Chen, E. H. 2014. Learning low-rank label correlations for multi-label classification with missing labels, *2014 IEEE international conference on data mining*, pp. 1067–1072.
- Xu, N., Liu, Y. P. and Geng, X. 2020. Partial multi-label learning with label distribution, *Proceedings of the AAAI Conference on Artificial Intelligence* **34**(4): 6510–6517.
- Yu, G. X., Chen, X., Domeniconi, C., Wang, J., Li, Z., Zhang, Z. L. and Wu, X. D. 2018. Feature-induced partial multi-label learning, *2018 IEEE international conference on data mining (ICDM)*, pp. 1398–1403.
- Zhan, W. and Zhang, M. L. 2017. Multi-label learning with label-specific features via clustering ensemble, *IEEE* pp. 129–136.
- Zhang, M. L. and Wu, L. 2014. Lift: Multi-label learning with label-specific features, *IEEE transactions on pattern analysis and machine intelligence* **37**(1): 107–120.
- Zhu, P. F., Xu, Q., Hu, Q. H., Zhang, C. Q. and Zhao, H. 2018. Multi-label feature selection with missing labels, *Pattern Recognition* **74**: 488–502.
- Zhu, Y., Kwok, J. T. and Zhou, Z. H. 2017. Multi-label learning with global and local label correlation, *IEEE Transactions on Knowledge and Data Engineering* **30**(6): 1081–1094.