Semantic-Aware Feature Enhancement for Partial Label Learning

Haowei Mei[1], Chao Zhang[1*], Wentao Fan[1], Xiuyi Jia[2], Chunlin Chen[1], Huaxiong Li[1]

[1]Nanjing University
[2]Nanjing University of Science and Technology
{hwmei, chzhang, fanwentao0955}@smail.nju.edu.cn, jiaxy@njust.edu.cn, {clchen, huaxiongli}@nju.edu.cn

Abstract

Partial label learning (PLL) aims to learn from the data where each instance is associated with a candidate label set, with only one being valid. Most existing approaches are designed to eliminate noisy labels and use the remaining reliable ones for model training, following a label-centric learning paradigm. In this paper, we propose a new PLL method called Semantic-Aware Feature Enhancement (SAFE), which tackles the problem through a novel feature-centric learning paradigm. SAFE presumes that the candidate labels are correct while the observed features are partial, and thus seeks to recover the underlying missing features. In this manner, a desired predictive model is constructed by integrating the observed and recovered features, which are responsible for predicting the true label and the remaining candidate labels, respectively. To ensure the quality of recovered features, SAFE jointly explores the intrinsic topological structures via dynamic graphs in both feature and label spaces as guidance for semantic-aware feature enhancement. Extensive experimental results on some popular datasets demonstrate the effectiveness and superiority of the proposed method over state-of-the-art PLL approaches.

Code — https://github.com/HHaoweimei/SAFE-PLL

Introduction

Partial label learning (PLL) represents a form of weakly supervised learning, where each training example is associated with a set of candidate labels, with exactly one being correct (Hüllermeier and Beringer 2006; Zhang, Zhou, and Liu 2016; Lyu et al. 2021; Tian, Yu, and Fu 2023). This learning paradigm has been successfully applied across various domains, including natural language processing (Zhang, Yu, and Tang 2017; Xu et al. 2021), web mining (Zhang et al. 2022; Jia, Yang, and Dong 2023; Gong, Yuan, and Bao 2022), and classification tasks (Zhang et al. 2025a).

The challenge of PLL lies in the ambiguity of label information. The only prior knowledge available is that the set of candidate labels definitely includes a true label. Thus, a natural way is label disambiguation, which aims to identify the potential true label via distinguishing the confidence of each candidate label and leverage the refined labels to guide the learning process. Representative label disambiguation techniques can be broadly categorized into two types: average-based methods (Zhang and Yu 2015; Cour et al. 2009) and identification-based methods (Si et al. 2024; Tang and Zhang 2017; Yan and Guo 2020). The former approach assigns identical importance to all candidate labels and derives the final prediction by averaging the model outputs corresponding to those labels (Zhou and Gu 2018; Gong et al. 2018; Liu and Dietterich 2014). Identification-based approaches regard the true label as a latent variable, with their primary objective being to uncover or infer this hidden variable. To achieve reliable disambiguation, various regularization techniques have been proposed, such as manifold regularization (Zhang et al. 2025b; Hou, Geng, and Zhang 2016; Jiang et al. 2025; Wang and Xu 2024; Zhang et al. 2024; Sun et al. 2025; Zhang et al. 2020), label distribution (Kou et al. 2025a; He, Feng, and Li 2018), label enhancement (Kou et al. 2025b; Xu, Lv, and Geng 2019; Sheng et al. 2024), and self-training (Feng and An 2019), etc. These various label disambiguation approaches focus on enhancing the accuracy of label information, and we refer to them as label-centric methods in this paper. Label disambiguation is an intuitive and natural way to address PLL; however, it may perform poorly when confronted with misleading candidate labels or insufficiently discriminative feature representations.

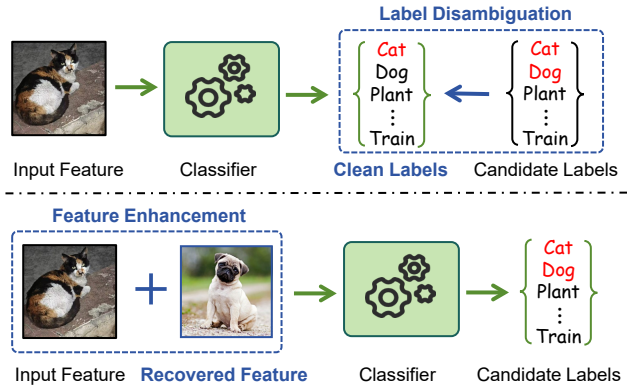


Figure 1: Comparison of label-centric learning paradigm (top) and our feature-centric learning paradigm (bottom).

*Corresponding author.

In this paper, we propose a novel feature-centric approach, Semantic-Aware Feature Enhancement (SAFE) for PLL. Different from label-centric approaches that treat incorrect candidate labels as noise, our method assumes that the observed candidate labels are correct, while the input features are partial or insufficient, and it aims to recover the missing features for trustworthy classification. Fig. 1 illustrates the difference between these two learning paradigms, which focus on the output and input space, respectively. The main contributions of this work are outlined as follows:

- We propose a novel feature-centric PLL method called SAFE. It assumes the label space is trustworthy and instead focuses on enhancing the feature representations of instances.
- The instance correlations within and across the feature and label spaces are fully exploited to enable reliable feature recovery and classifier construction.
- Extensive experiments are conducted on popular PLL datasets, and the results validate the effectiveness and superiority of SAFE over existing methods.

Related Work

PLL is a typical weakly supervised learning paradigm that operates under ambiguous supervisory information. One popular technique to solve it is label disambiguation (Lu et al. 2025; Jia et al. 2019), which can generally be categorized into two classes: average-based methods and identification-based methods.

As a representative average-based method, PL-KNN (Hüllermeier and Beringer 2006) makes predictions by learning from ambiguously labeled examples, using a k-nearest neighbors approach to infer class labels based on the most consistent label assignments among neighboring instances. This strategy is intuitive; however, it becomes less effective when the candidate label set is predominantly composed of noisy and ambiguous labels. In identification-based methods (Xu, Lv, and Geng 2019; Lyu, Wu, and Feng 2022; Chen et al. 2020), the true label is treated as a latent variable that can be inferred through an iterative optimization procedure. For instance, PL-AGGD (Wang, Zhang, and Li 2022) classifies instances by adaptively constructing a graph that guides the disambiguation of partially labeled data, refining label predictions through the relationships among instances. SURE (Feng and An 2019) leverages self-training to deal with the challenge of ambiguous labels, using maximum infinity norm regularization to automatically distinguish high-confidence candidate labels as true labels. PL-CL (Jia, Si, and Zhang 2023) leverages both candidate and non-candidate labels to build a complementary classifier, reducing false positives and enhancing label disambiguation. NLR (Yang et al. 2024) eliminates noisy labels using prior knowledge and a competitive learning model. With deep learning advances, methods (Lv et al. 2020; Wu, Wang, and Zhang 2022; He et al. 2022; Lyu, Wu, and Feng 2022) employ neural networks for discriminative feature learning and improved label disambiguation.

Despite employing different strategies, these previous methods generally solve the PLL problem from a label-centric perspective, treating candidate labels as corrupted and aiming to identify the true label by filtering out the incorrect ones. However, it may be difficult to accurately identify the ground-truth label, especially when noisy labels closely resemble the true label or dominate the candidate set. In this paper, we propose a new PLL method based on semantic-aware feature enhancement, marking a fundamental shift toward a feature-centric learning paradigm.

The Proposed Approach

The goal of PLL is to induce a classifier $f : \mathcal{X} \to \mathcal{Y}$ based on the partial labeled dataset. Let $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n]^\mathsf{T} \in \mathbb{R}^{n \times q}$ denote the training instance matrix and $\mathbf{Y} = [\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_n]^\mathsf{T} \in \{0, 1\}^{n \times l}$ denote the partial label matrix, where n, l, and q represent the number of instances, classes and feature dimension, respectively. $\mathbf{Y}$ is a binary matrix, and $y_{ij} = 1$ (or $y_{ij} = 0$) indicates that the j-th label is present in (or absent from) the candidate label set of $\mathbf{x}_i$. $\mathbf{1}_n$ denotes an n-dimensional column vector of ones, $\mathbf{1}_{n \times l}$ denotes an $n \times l$ all-ones matrix, and $\mathbf{0}_{n \times q}$ denotes an $n \times q$ zero matrix.

Full Label Prediction with Feature Enhancement

Previous label-centric methods typically regard candidate labels as noisy or corrupted, and they focus on disambiguating them to recover the latent true label. In contrast, our SAFE approach starts from a different perspective: it assumes that the provided candidate labels are correct, while the observed features are partial or incomplete. Under this assumption, SAFE shifts the focus from label correction to feature completion, aiming to recover the missing feature information to construct a trustworthy classifier. This can be achieved by solving the following model:

$$\min_{\hat{\mathbf{X}}, \mathbf{b}, \mathbf{W}} \quad \left\| \left(\mathbf{X} + \hat{\mathbf{X}} \right) \mathbf{W} + \mathbf{2}_n \mathbf{b}^\mathsf{T} - \mathbf{Y} \right\|_F^2 + \lambda \left\| \mathbf{W} \right\|_F^2, \quad (1)$$

where $\hat{\mathbf{X}} = [\hat{\mathbf{x}}_1, \hat{\mathbf{x}}_2, \ldots, \hat{\mathbf{x}}_n]^\mathsf{T}$ represents the recovered missing features, $\mathbf{b} \in \mathbb{R}^l$ is a bias vector, and λ is a regularization parameter. Eq. (1) jointly recovers the missing features and learns a classifier in a unified framework to predict the full candidate labels $\mathbf{Y}$. $\mathbf{x}_i + \hat{\mathbf{x}}_i$ is an enhanced instance feature vector that contains sufficient semantic information corresponding to its candidate label $\mathbf{y}_i$.

One prior in PLL is that the candidate label $\mathbf{y}_i$ contains a true label of $\mathbf{x}_i$. To leverage this prior, we extend Eq. (1) to the following formulation:

$$\begin{aligned} \min_{\hat{\mathbf{X}}, \mathbf{b}, \mathbf{W}, \mathbf{P}} \quad & \left\| \left(\mathbf{X} + \hat{\mathbf{X}} \right) \mathbf{W} + \mathbf{2}_n \mathbf{b}^\mathsf{T} - \mathbf{Y} \right\|_F^2 + \lambda \left\| \mathbf{W} \right\|_F^2 \\ & + \alpha \left\| \mathbf{X}\mathbf{W} + \mathbf{1}_n \mathbf{b}^\mathsf{T} - \mathbf{P} \right\|_F^2 \\ \text{s.t.} \quad & \mathbf{P}\mathbf{1}_l = \mathbf{1}_n, \mathbf{0} \leq \mathbf{P} \leq \mathbf{Y}, \end{aligned} \quad (2)$$

where $\mathbf{P} \in \mathbb{R}^{n \times l}$ denotes the confidence distribution over labels for original input feature $\mathbf{x}_i$. The constraints on $\mathbf{P}$ encourage the model to adaptively identify the underlying true label from all candidates. In this manner, the original

features are used to predict the true label by $\mathbf{X}\mathbf{W} + \mathbf{1}_n\mathbf{b}^\top$, while the recovered features predict the remaining candidate labels by $\hat{\mathbf{X}}\mathbf{W} + \mathbf{1}_n\mathbf{b}^\top$.

Dual Topological Structure Embedding

Eq. (2) links the feature and label spaces by capturing feature–label associations. To improve robustness and feature recovery, we also leverage instance correlations through the topological structures of both spaces. Although the feature and label spaces lie on different manifolds, their local structures are similar. Specifically, if $\mathbf{x}_i$ and $\mathbf{x}_j$ are similar in feature space, their corresponding label confidence vectors $\mathbf{p}_i$ and $\mathbf{p}_j$ should also be similar. To this end, we construct a similarity matrix to capture the local structure in the feature space, and use it to guide the label confidence learning in the label space. Defining a kNN matrix $\mathbf{U} = [u_{ij}]_{n\times n}$, where $u_{ij} = 1$ if $\mathbf{x}_i$ is among the kNN of $\mathbf{x}_j$ and $u_{ij} = 0$ otherwise, we embed it into label space by

$$\begin{aligned} \min_{\mathbf{S}} \quad & \left\|\mathbf{P}^\top - \mathbf{P}^\top\mathbf{S}\right\|_F^2 \\ \text{s.t.} \quad & \mathbf{S}^\top\mathbf{1}_n = \mathbf{1}_n, \mathbf{0}_{n\times n} \le \mathbf{S} \le \mathbf{U}, \end{aligned} \tag{3}$$

where $\mathbf{S} = [s_{ij}]_{n\times n}$ is a learnable similarity matrix. The constraint $\mathbf{S}^\top\mathbf{1}_n = \mathbf{1}n$ normalizes $\mathbf{S}$, while $\mathbf{0}n \times n \le \mathbf{S} \le \mathbf{U}$ incorporates the feature-space local structure, aligning $\mathbf{P}$ with similar topologies. Eq. (3) further learns $\mathbf{S}$ adaptively, mitigating noise in the predefined $\mathbf{U}$.

The focus of Eq. (3) is to transfer the local structure in the feature space into label space, which improves the quality of $\mathbf{P}$. To further enhance the quality of the recovered features, we simultaneously embed the local structure from the label space back into the feature space. Defining a kNN semantic similar matrix $\mathbf{V} = [v_{ij}]_{n\times n}$, where $v_{ij} = 1$ if $\mathbf{y}_i$ is among the kNN of $\mathbf{y}_j$ based on cosine similarity, and $v_{ij} = 0$ otherwise, we then embed it into feature space by

$$\begin{aligned} \min_{\mathbf{M}} \quad & \left\|\hat{\mathbf{X}}^\top - \hat{\mathbf{X}}^\top\mathbf{M}\right\|_F^2 \\ \text{s.t.} \quad & \mathbf{M}^\top\mathbf{1}_n = \mathbf{1}_n, \mathbf{0}_{n\times n} \le \mathbf{M} \le \mathbf{V}, \end{aligned} \tag{4}$$

where $\mathbf{M} = [m_{ij}]_{n\times n}$ is also a learnable similarity matrix. In PLL, the semantic similarity matrix $\mathbf{V}$ contains false positive entries for recovered feature $\hat{\mathbf{X}}$, i.e., false "1". By adaptively learning a similarity matrix $\mathbf{M}$, the negative influence of those noisy information can be reduced.

Overall Objective Formulation

The overall objective of SAFE is formulated as follows:

$$\begin{aligned} \min_{\Omega} \; & \left\|\left(\mathbf{X} + \hat{\mathbf{X}}\right)\mathbf{W} + \mathbf{2}_n\mathbf{b}^\top - \mathbf{Y}\right\|_F^2 + \lambda\left\|\mathbf{W}\right\|_F^2 \\ & + \beta\left\|\mathbf{P}^\top - \mathbf{P}^\top\mathbf{S}\right\|_F^2 + \gamma\left\|\hat{\mathbf{X}}^\top - \hat{\mathbf{X}}^\top\mathbf{M}\right\|_F^2 \\ & + \alpha\left\|\mathbf{X}\mathbf{W} + \mathbf{1}_n\mathbf{b}^\top - \mathbf{P}\right\|_F^2 \\ \text{s.t.} \; & \mathbf{P}\mathbf{1}_l = \mathbf{1}_n, \mathbf{0}_{n\times l} \le \mathbf{P} \le \mathbf{Y}, \mathbf{S}^\top\mathbf{1}_n = \mathbf{1}_n, \\ & \mathbf{0}_{n\times n} \le \mathbf{S} \le \mathbf{U}, \mathbf{M}^\top\mathbf{1}_n = \mathbf{1}_n,\ \mathbf{0}_{n\times n} \le \mathbf{M} \le \mathbf{V}, \end{aligned} \tag{5}$$

where $\Omega = \{\hat{\mathbf{X}}, \mathbf{b}, \mathbf{W}, \mathbf{P}, \mathbf{S}, \mathbf{M}\}$ is the variable set, α, β, γ and λ are balance parameters. The above model addresses the PLL problem from the perspective of feature completion. By jointly embedding the local structures between the feature and label spaces, it can effectively exploit complementary information to construct a trustworthy classifier.

Optimization

The optimization problem in Eq. (5) involves six variables with different constraints. We adopt an alternating and iterative approach to solve it.

Update $\mathbf{W}$, $\mathbf{b}$, and $\hat{\mathbf{X}}$ With all other variables fixed, the problem with respect to $\mathbf{W}$, $\mathbf{b}$, and $\hat{\mathbf{X}}$ is reformulated as:

$$\begin{aligned} \min_{\mathbf{W},\mathbf{b},\hat{\mathbf{X}}} \; & \left\|\left(\mathbf{X} + \hat{\mathbf{X}}\right)\mathbf{W} + \mathbf{2}_n\mathbf{b}^\top - \mathbf{Y}\right\|_F^2 + \gamma\left\|\hat{\mathbf{X}}^\top - \hat{\mathbf{X}}^\top\mathbf{M}\right\|_F^2 \\ & + \alpha\left\|\mathbf{X}\mathbf{W} + \mathbf{1}_n\mathbf{b}^\top - \mathbf{P}\right\|_F^2 + \lambda\left\|\mathbf{W}\right\|_F^2. \end{aligned} \tag{6}$$

This is a regularized least-squares problem, and a closed-form solution can be obtained by taking derivatives and rearranging terms. We obtain:

$$\begin{aligned} \mathbf{W} = & \left(\mathbf{A}^\top\mathbf{A} + \alpha\mathbf{X}^\top\mathbf{X} + \lambda\mathbf{I}_{n\times n}\right)^{-1} \\ & \cdot\left(\mathbf{A}^\top\mathbf{Y} + \alpha\mathbf{X}^\top\mathbf{P} - \left(\alpha\mathbf{X} + 2\mathbf{A}\right)^\top\mathbf{1}_n\mathbf{b}^\top\right), \end{aligned} \tag{7}$$

where $\mathbf{A} = \mathbf{X} + \hat{\mathbf{X}}$, $\mathbf{I}_{n\times n}$ is an $n \times n$ identity matrix. Similarly, we obtain:

$$\mathbf{b} = \frac{1}{(4+\alpha)n}\left(\alpha\mathbf{P}^\top + 2\mathbf{Y}^\top - \mathbf{W}^\top\left(\alpha\mathbf{X}^\top + 2\mathbf{A}^\top\right)\right)\mathbf{1}_n, \tag{8}$$

Updating $\hat{\mathbf{X}}$ reduces to solving a Sylvester equation:

$$\gamma\mathbf{H}\hat{\mathbf{X}} + \hat{\mathbf{X}}\mathbf{W}\mathbf{W}^\top = \left(\mathbf{Y} - \mathbf{2}_n\mathbf{b}^\top - \mathbf{X}\mathbf{W}\right)\mathbf{W}^\top, \tag{9}$$

where $\mathbf{H} = (\mathbf{I}_{n\times n} - \mathbf{M}^\top)^\top(\mathbf{I}_{n\times n} - \mathbf{M}^\top)$. This standard Sylvester equation can be solved with numerical solvers.

Kernel Transformation The above linear model may be insufficient for modeling complex data. To better capture the intricate and nonlinear relationships between instances and labels, we introduce a kernel transformation that projects the original feature space into a higher-dimensional Hilbert space (Jia, Yang, and Dong 2023). Let $\phi(\cdot) : \mathbb{R}^q \to \mathbb{R}^h$ be a feature mapping that transforms the original feature space into a higher-dimensional space. In this paper, we adopt the Gaussian kernel function, and use the kernelized feature $\mathbf{\Phi}(\mathbf{X}) = [\phi(\mathbf{x_1}), \phi(\mathbf{x}_2), \cdots, \phi(\mathbf{x})_n]^\top$ for training.

Update $\mathbf{S}$ and $\mathbf{M}$ With all other variables fixed, the $\mathbf{S}$ subproblem can be rewritten as:

$$\begin{aligned} \min_{\mathbf{S}} \quad & \left\|\mathbf{P}^\top - \mathbf{P}^\top\mathbf{S}\right\|_F^2 \\ \text{s.t.} \quad & \mathbf{S}^\top\mathbf{1}_n = \mathbf{1}_n,\ \mathbf{0}_{n\times n} \le \mathbf{S} \le \mathbf{U}. \end{aligned} \tag{10}$$

The columns of $\mathbf{S}$ are independent, allowing the subproblem to be solved in a column-wise manner. For the i-th column, the resulting subproblem is:

$$\begin{aligned} \min_{\mathbf{S}_{\cdot i}} \quad & \left\|\mathbf{p}_i - \sum_{k_{ji}=1} s_{ji}\mathbf{p}_j\right\|_F^2 \\ \text{s.t.} \quad & \mathbf{S}_{\cdot i}^\top\mathbf{1}_n = 1, \mathbf{0}_n \le \mathbf{S}_{\cdot i} \le \mathbf{U}_{\cdot i}. \end{aligned} \tag{11}$$

Given that only k elements in $\mathbf{S}_{\cdot i}$ are non-zero, these elements correspond to the coefficients used for reconstructing $\mathbf{x}_i$ from its nearest neighbors. Denote the vector composed of these coefficients as $\hat{\mathbf{s}}_i \in \mathbb{R}^k$. Let $\mathcal{N}_i$ denote the index set associated with these neighbors. Then, we introduce the following matrices: $\mathbf{Q}^{p_i} = [\mathbf{p}_i - \mathbf{p}_{\mathcal{N}_{i(1)}}, \mathbf{p}_i - \mathbf{p}_{\mathcal{N}_{i(2)}}, \ldots, \mathbf{p}_i - \mathbf{p}_{\mathcal{N}_{i(k)}}]^\mathsf{T} \in \mathbb{R}^{k \times l}$, and define the Gram matrices: $\mathbf{G}^{p_i} = \mathbf{Q}^{p_i}(\mathbf{Q}^{p_i})^\mathsf{T} \in \mathbb{R}^{k \times k}$.Then, Eq. (11) can be reformulated as:

$$\begin{aligned} \min_{\hat{\mathbf{s}}_i} \quad & \hat{\mathbf{s}}_i^\mathsf{T} \mathbf{G}^{p_i} \hat{\mathbf{s}}_i \\ \text{s.t.} \quad & \hat{\mathbf{s}}_i^\mathsf{T} \mathbf{1}_k = 1, \mathbf{0}_k \leq \hat{\mathbf{s}}_i \leq \mathbf{1}_k. \end{aligned} \tag{12}$$

The optimization problem in Eq. (12) constitutes a standard quadratic programming (QP) problem, which can be addressed using conventional QP solvers. Finally, the matrix $\mathbf{S}$ is constructed by concatenating all the obtained $\hat{\mathbf{s}}_i$ vectors.

The sub-problem w.r.t. $\mathbf{M}$ is the same with $\mathbf{S}$, and we can use the same strategy to solve it.

Update $\mathbf{P}$ With all other variables fixed, the $\mathbf{P}$ subproblem is formulated as:

$$\begin{aligned} \min_{\mathbf{P}} \quad & \alpha \left\| \mathbf{X}\mathbf{W} + \mathbf{1}_n \mathbf{b}^\mathsf{T} - \mathbf{P} \right\|_F^2 + \beta \left\| \mathbf{P}^\mathsf{T} - \mathbf{P}^\mathsf{T}\mathbf{S} \right\|_F^2 \\ \text{s.t.} \quad & \mathbf{P}\mathbf{1}_l = \mathbf{1}_n, \mathbf{0}_{n \times l} \leq \mathbf{P} \leq \mathbf{Y}. \end{aligned} \tag{13}$$

We rewrite Eq. (13) in the following form:

$$\begin{aligned} \min_{\mathbf{P}} \quad & \left\| \mathbf{P} - (\mathbf{X}\mathbf{W} + \mathbf{1}_n \mathbf{b}^\mathsf{T}) \right\|_F^2 + \frac{\beta}{\alpha} \left\| \mathbf{P}^\mathsf{T} - \mathbf{P}^\mathsf{T}\mathbf{S} \right\|_F^2 \\ \text{s.t.} \quad & \mathbf{P}\mathbf{1}_l = \mathbf{1}_n, \mathbf{0}_{n \times l} \leq \mathbf{P} \leq \mathbf{Y}. \end{aligned} \tag{14}$$

To address problem (14), we first introduce the notation $\widetilde{\mathbf{p}} = \mathrm{vec}(\mathbf{P}) \in [0,1]^{nl}$, where $\mathrm{vec}(\mathbf{P})$ denotes the vectorization operator. Likewise, we define $\widetilde{\mathbf{o}} = \mathrm{vec}(\mathbf{X}\mathbf{W} + \mathbf{1}_n \mathbf{b}^\mathsf{T}) \in \mathbb{R}^{nl}$ and $\widetilde{\mathbf{y}} = \mathrm{vec}(\mathbf{Y}) \in \{0,1\}^{nl}$. Let $\mathbf{T} = 2(\mathbf{I}_{n \times n} - \mathbf{S})(\mathbf{I}_{n \times n} - \mathbf{S})^\mathsf{T} \in \mathbb{R}^{n \times n}$ denote a square matrix. Under these definitions, the optimization problem in (14) can be equivalently rewritten as follows:

$$\begin{aligned} \min_{\widetilde{\mathbf{p}}} \quad & \frac{1}{2} \widetilde{\mathbf{p}}^\mathsf{T} \left(\mathbf{E} + \frac{2\alpha}{\beta} \mathbf{I}_{nl \times nl} \right) \widetilde{\mathbf{p}} - \frac{2\alpha}{\beta} \widetilde{\mathbf{o}}^\mathsf{T} \widetilde{\mathbf{p}} \\ \text{s.t.} \quad & \sum_{i=1,\, i\%n=j,\, 0 \leq j \leq n-1}^{nl} \widetilde{\mathbf{p}}_i = 1, \mathbf{0}_{nl} \leq \widetilde{\mathbf{p}} \leq \widetilde{\mathbf{y}}, \end{aligned} \tag{15}$$

where $\mathbf{E} \in \mathbb{R}^{nl \times nl}$ is defined as:

$$\mathbf{E} = \begin{bmatrix} \mathbf{T} & \mathbf{0}_{m \times m} & \cdots & \mathbf{0}_{m \times m} \\ \mathbf{0}_{m \times m} & \mathbf{T} & \ddots & \vdots \\ \vdots & \ddots & \ddots & \mathbf{0}_{m \times m} \\ \mathbf{0}_{m \times m} & \cdots & \mathbf{0}_{m \times m} & \mathbf{T} \end{bmatrix}. \tag{16}$$

Eq. (15) is also a standard QP problem, and it can be solved using existing QP solvers.

After solving the model (5), the projection matrix $\mathbf{W}$ and bias $\mathbf{b}$ can be obtained. Then, for an unseen instance $\mathbf{x}^*$, its label y^* is predicted by

$$y^* = \arg\max_k \left(\mathbf{\Phi}(\mathbf{x}^*) \mathbf{W}_{\cdot k} + b_k \right). \tag{17}$$

The pseudo code of SAFE is summarized in Algorithm 1.

Algorithm 1: SAFE algorithm

Require: Data $\{\mathbf{X}, \mathbf{Y}\}$, hyper-parameters α, β, γ, and λ.
Ensure: $\mathbf{W}$ and $\mathbf{b}$.

```
Construct the kernel matrix Φ(X).
Initialize S = M = I_{n×n}, P = Y, W = 1_{n×l}, b = 1_l, and X̂ = 0_{n×q}.
repeat
    Update W, b, X̂ via Eqs. (7), (8), and (9);
    Update S, M via Eq. (12);
    Update P via Eq. (15);
until Convergence
Return W, b.
```

Complexity Analysis

The complexity of solving for $\mathbf{W}$, $\mathbf{b}$, and $\hat{\mathbf{X}}$ is $O(n^3)$, $O(nql)$, and $O(n^3)$, respectively. The update of $\mathbf{S}$ and $\mathbf{M}$ has a complexity of $O(nk^3)$ and $O(nk^3)$, while the update of $\mathbf{P}$ has a complexity of $O(n^3q^3)$. Therefore, the overall complexity of SAFE is $O(2n^3 + nql + 2nk^3 + n^3q^3)$.

Experiments

Experimental Setup

Datasets We conduct experiments on six real-world PLL datasets, including Lost (Cour et al. 2009), FG-NET (Panis et al. 2016), MSRCv2 (Liu and Dietterich 2012), Mirflickr (Huiskes and Lew 2008), Soccer Player (Zeng et al. 2013), and Yahoo! News (Guillaumin, Verbeek, and Schmid 2010), as well as four UCI datasets, including ecoli, abalone, vehicle, and segment. Table 1 summarizes dataset statistics.

Baselines To validate the effectiveness of SAFE, we compared it with the following approaches:

- NLR (Yang et al. 2024): Constructs a competitive learning model to identify and remove noisy labels for label disambiguation. [$\lambda \in [1,4]$ with step size 0.1, β and $\gamma \in \{0.001, 0.01, 0.1, 1, 10, 100, 1000\}$].
- PL-CL (Jia, Si, and Zhang 2023): Constructs a complementary classifier and a similarity graph for label disambiguation. [$k = 10$; $\lambda = 0.03$, γ, μ, α, $\beta \in \{0.001, 0.01, 0.1, 0.2, 0.5, 1, 1.5, 2, 4\}$].
- PL-AGGD (Wang, Zhang, and Li 2022): Uses similarity graphs to guide disambiguation. [$k = 10$, $T = 20$, $\lambda = 1$, $\mu = 1$, $\gamma = 0.05$].
- SURE (Feng and An 2019): Employs self-guided retraining with infinity norm regularization for label filtering. [$\lambda, \beta \in \{0.001, 0.01, 0.05, 0.1, 0.3, 0.5, 1\}$].
- LALO (Feng and An 2018): Applies local consistency to separate true and false labels. [$k = 10$, $\lambda = 0.05$, $\mu = 0.005$].
- IPAL (Zhang and Yu 2015): Performs iterative label propagation to refine candidate labels. [$\alpha = 0.95$, $k = 10$, $T = 100$].
- PLDA (Wang and Zhang 2022): Integrates dimensionality reduction with PLL. [$k = 10$].

Datasets	Real-world datasets						UCI datasets			
	Lost	FG-NET	MSRCv2	Mirflickr	Soccer Player	Yahoo!News	ecoli	abalone	vehicle	segment
# Instances	1122	1002	1758	2780	17472	22991	336	4177	846	2310
# Features	108	262	48	1536	279	163	7	7	18	19
# Labels	16	78	23	14	171	219	8	29	4	7
# Average Labels	2.23	7.48	3.16	2.76	2.09	1.91	-	-	-	-

Table 1: General characteristics of the real-world and UCI datasets.

Method	FG-NET	Lost	MSRCv2	Mirflickr	Soccer Player	Yahoo!News	FG3	FG5
SAFE	**0.070±0.010**	**0.724±0.015**	**0.478±0.018**	**0.666±0.010**	0.539±0.002	0.611±0.003	**0.426±0.014**	**0.574±0.008**
NLR	0.067±0.009	0.621±0.024•	0.465±0.014	0.613±0.018•	0.489±0.003•	0.568±0.007•	0.403±0.016•	0.553±0.014•
PL-CL	0.068±0.008	0.713±0.017	0.469±0.011	0.663±0.009	0.534±0.004•	**0.614±0.002**	0.423±0.018	0.565±0.010
SURE	0.064±0.007	0.695±0.023•	0.464±0.021	0.645±0.008•	**0.548±0.002**○	0.602±0.004	0.401±0.018•	0.567±0.014
PL-AGGD	0.066±0.009	0.700±0.013•	0.465±0.012	0.612±0.011•	0.527±0.004•	0.609±0.004	0.393±0.015•	0.561±0.012•
LALO	0.066±0.007	0.714±0.020	0.465±0.017	0.659±0.004•	0.528±0.003•	0.606±0.003•	0.408±0.018•	0.563±0.011•
IPAL	0.050±0.012•	0.572±0.026•	0.462±0.012•	0.551±0.017•	0.528±0.003•	0.564±0.004•	0.352±0.020•	0.512±0.019•
PLDA	0.043±0.011•	0.381±0.020•	0.390±0.016•	0.482±0.016•	0.495±0.001•	0.450±0.004•	0.343±0.019•	0.511±0.019•

Table 2: Classification accuracy of different methods on real-world datasets. •/○ indicates whether SAFE is statistically superior/inferior to the compared algorithm according to pairwise t-test at significance level of 0.05. Bold indicates the best results. FG3 and FG5 denote FG-NET(MAE3) and FG-NET(MAE5).

	I	**II**	**III**	**IV**	**Total**
NLR	23/5/0	28/0/0	20/6/2	25/3/0	102/10/0
PL-CL	16/10/2	18/10/0	16/12/0	17/11/0	67/43/2
SURE	18/10/0	15/13/0	14/14/0	18/9/1	65/46/1
PL-AGGD	17/11/0	19/9/0	19/9/0	15/13/0	70/42/0
LALO	20/7/1	17/11/0	12/16/0	14/14/0	63/48/1
IPAL	24/3/1	21/1/6	21/4/3	23/4/1	89/12/11
PLDA	28/0/0	28/0/0	28/0/0	28/0/0	112/0/0

Table 3: Win/tie/loss counts on the controlled UCI data sets between SAFE and other compared approaches based on direct value comparison. I: varying ϵ ($p = 1, r = 1$); II: varying p ($r = 1$); III: varying p ($r = 2$); IV: varying p ($r = 3$).

Implementation Details For our SAFE, we set $k = 10$, $\lambda = 0.03$. The parameters α, β, and γ were selected from the set $\{0.0001, 0.001, 0.01, 0.03, 0.1, 0.5, 1, 5, 10, 30\}$. The baselines NLR, PL-CL, SURE, PL-AGGD, and LALO, together with our proposed SAFE, all employ the Gaussian kernel function. All comparison algorithms are evaluated on five random splits of the dataset into 50%/50% training and test sets, and the average accuracy along with the standard deviation are reported.

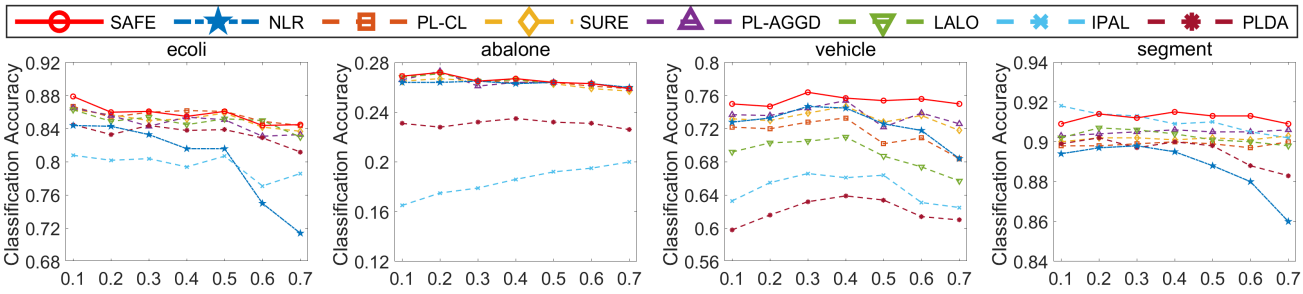
Performance on Real-World Datasets

Table 2 presents the classification accuracy along with the corresponding standard deviation for each method evaluated on real-world datasets. FG-NET is a facial age estimation dataset, in which age annotations collected from crowd-sourced sources are treated as candidate labels. The average number of candidate labels in FG-NET is relatively large, which may result in low classification accuracy under conventional evaluation metrics. To better assess performance on this task, we adopt the mean absolute error (MAE) (Zhang, Zhou, and Liu 2016; Jia, Si, and Zhang 2023) and further define two variants, FG-NET (MAE3) and FG-NET (MAE5), where a test sample is deemed correctly classified if the predicted age differs from the ground-truth age by at most 3 or 5 years, respectively. As illustrated in Table 2, it can be observed that:

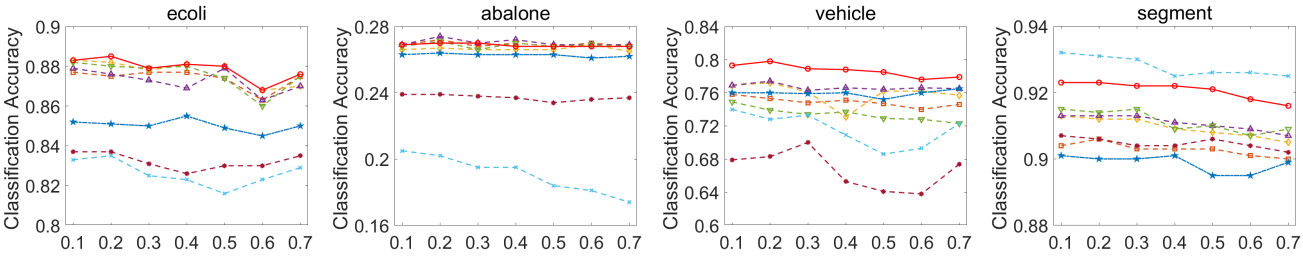
- SAFE consistently outperforms NLR, PL-AGGD, LALO, IPAL, and PLDA methods on six real-world datasets (Lost, FG-NET, FG3, FG5, Mirflickr, and MSRCv2). It also achieves superior performance compared to graph-based approaches such as PL-AGGD and PL-CL, validating the effectiveness of its semantic-aware feature enhancement strategy.
- While SURE and PL-CL achieve the best results on the Soccer Player and Yahoo!News datasets, respectively, their performance is inferior to that of our SAFE method on the remaining datasets. Moreover, SAFE delivers competitive performance on both the Soccer Player and Yahoo!News datasets, outperforming other baselines and ranking second.
- Overall, according to pairwise t-test at a significance level of 0.05, SAFE achieves significantly higher accuracy in 64.3% of cases and is significantly outperformed in only 1.8% of cases.

Performance on Controlled UCI Datasets

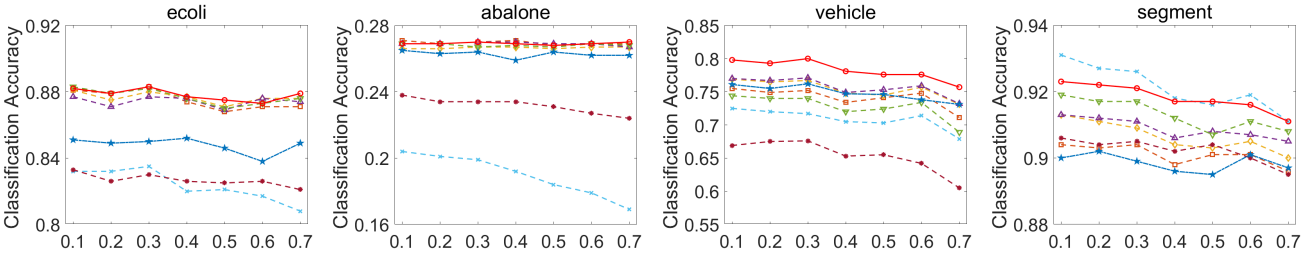
To further evaluate the robustness of the model, we conduct experiments on four UCI datasets under controlled label settings. Following (Zeng et al. 2013; Liu and Dietterich 2012; Jia, Si, and Zhang 2023), we generate synthetic datasets with partial labels by adjusting three parameters: p represents the proportion of training instances associated with partial la-



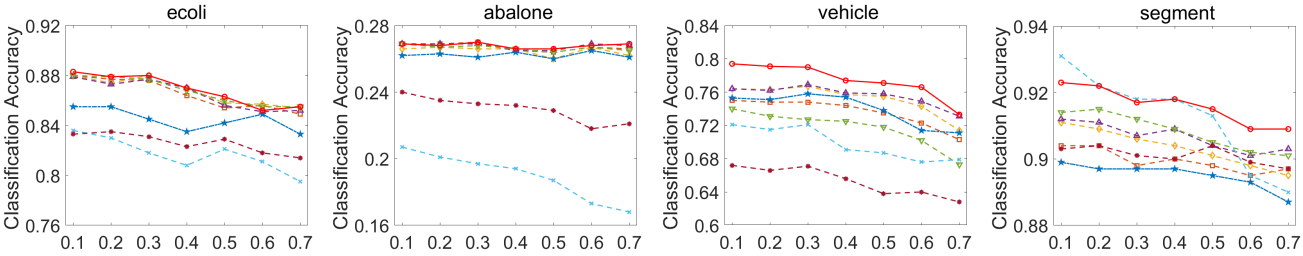
(a) Classification accuracy on controlled UCI datasets with ϵ varying from 0.1 to 0.7 ($p = 1$, $r = 1$).



(b) Classification accuracy on controlled UCI datasets with p varying from 0.1 to 0.7 ($r = 1$).



(c) Classification accuracy on controlled UCI datasets with p varying from 0.1 to 0.7 ($r = 2$).



(d) Classification accuracy on controlled UCI datasets with p varying from 0.1 to 0.7 ($r = 3$).

Figure 2: Classification accuracy on the controlled UCI data sets under different settings.

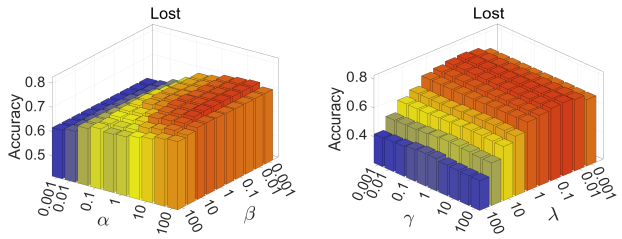
bels; r denotes the number of false positive labels included in each candidate set; and ϵ indicates the likelihood that an incorrect label co-occurs with the true label.

Fig. 2(a) reports the classification results as ϵ increases from 0.1 to 0.7, with $p = r = 1$ fixed. A specific label is selected to co-occur with the ground-truth label with probability ϵ, while the remaining incorrect labels are randomly sampled with probability $(1 - \epsilon)/(l - 2)$. Figs. 2(b)–(d) present the classification results under varying p from 0.1 to 0.7, with $r = 1$, 2, and 3, respectively. In these settings, a proportion p of the training instances are treated as partial label samples, each augmented with r randomly selected false positive labels, resulting in candidate sets of size $r+1$. Table 3 reports the win/tie/loss statistics between SAFE and other baseline methods in the four settings, based on pairwise t-test at a significance level of 0.05. From the Fig. 2 and Table 3, we can observe that:

- As p and ϵ increase, classification accuracy generally decreases for all methods, as larger values of p or ϵ indicate a more challenging PLL task. Nevertheless, our approach still outperforms most baselines in the majority of cases.
- As the number of noisy candidate labels increases, the performance advantages of our SAFE method over the compared methods become more obvious. Comparing the settings II and IV in Table 3, SAFE wins more cases in setting IV, which indicates that our method exhibits

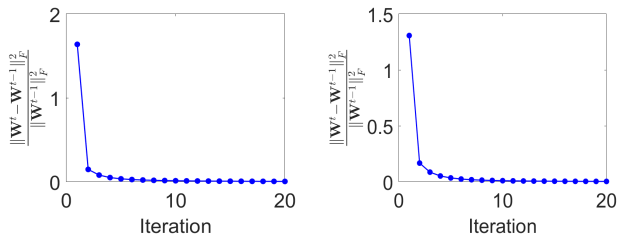
Kernel	FE	DTSE	FG-NET	Lost	MSRCv2	Mirflickr	Soccer Player	Yahoo!News	FG3	FG5
✗	✗	✗	0.065 ± 0.007	0.648 ± 0.028	0.381 ± 0.010	0.476 ± 0.021	0.493 ± 0.002	0.455 ± 0.004	0.398 ± 0.017	0.553 ± 0.015
✓	✗	✗	0.065 ± 0.007	0.667 ± 0.020	0.453 ± 0.012	0.634 ± 0.010	0.503 ± 0.003	0.563 ± 0.003	0.420 ± 0.020	0.565 ± 0.009
✓	✗	✓	0.066 ± 0.011	0.684 ± 0.011	0.441 ± 0.022	0.665 ± 0.013	0.500 ± 0.003	0.559 ± 0.006	0.424 ± 0.024	0.566 ± 0.013
✓	✓	✗	0.067 ± 0.010	0.719 ± 0.014	0.471 ± 0.017	0.660 ± 0.013	0.536 ± 0.002	0.606 ± 0.006	0.411 ± 0.015	0.564 ± 0.012
✓	✓	✓	**0.070 ± 0.010**	**0.724 ± 0.015**	**0.478 ± 0.018**	**0.666 ± 0.010**	**0.539 ± 0.002**	**0.611 ± 0.003**	**0.426 ± 0.014**	**0.574 ± 0.008**

Table 4: Ablation study results on real-world datasets. Bold indicates the best performing results. FE and DTSE denote feature enhancement and dual topology structure embedding. FG3 and FG5 denote FG-NET(MAE3) and FG-NET(MAE5).



(a) α and β (b) γ and λ

Figure 3: Parameter sensitivity analysis on Lost dataset.



(a) Lost (b) MSRCv2

Figure 4: Convergence curves on two datasets.

greater robustness against noisy candidate labels.

- Generally speaking, SAFE clearly outperforms other methods under most settings. Specifically, SAFE outperforms other methods significantly in 72.4% of cases and is only significantly outperformed in 1.9% of cases.

Ablation Study

To validate the effectiveness of feature enhancement (FE) and dual topological structure embedding (DTSE) in SAFE, we conduct ablation studies on six real-world datasets, and report the results in Table 4. The role of kernel transformation is also evaluated. (1) Comparing Row 1 with Row 2, we observe that kernel transformation enhances nonlinear modeling capability, leading to improved classification performance, particularly on the MSRCv2, Mirflickr, and Yahoo!News datasets. (2) Comparing Row 2 with Row 4, we find that feature enhancement further improves accuracy, especially on Lost, Mirflickr, Soccer Player, and Yahoo!News, validating the effectiveness of feature recovery. (3) Comparing the last two rows, we observe that dual topological structure embedding explores local structures within both the feature and label spaces, which enhances model robustness and overall performance. Compared with all variants, the proposed SAFE model achieves the best results across all datasets, demonstrating the effectiveness of both feature enhancement and dual topological structure embedding.

Parameter Sensitivity Analysis

Our SAFE model involves four hyper-parameters, α, β, γ, and λ, which influence its classification performance. To analyze the sensitivity of these parameters, we define a candidate range for each parameter and record the classification accuracy under different parameter combinations. Fig. 3 shows the classification results of SAFE w.r.t. the four hyper-parameters on the Lost dataset. During the analysis of any two parameters, the others are kept fixed. As the values of these parameters change, the classification accuracy of the SAFE model also fluctuates. Compared with β and λ, the classification performance shows more sensitivity to α and γ. Overall, SAFE maintains stable performance across a reasonably wide range of hyper-parameter values.

Convergence Analysis

To investigate the convergence property of the optimization algorithm, we define the stopping criterion based on the change rate of $\mathbf{W}$. Specifically, we define the loss value $\|\mathbf{W}^t - \mathbf{W}^{t-1}\|_F^2 / \|\mathbf{W}^{t-1}\|_F^2$. Fig. 4 shows the evolution of this value w.r.t. the number of iterations on the Lost and MSRCv2 datasets. As can be observed, the loss value decreases rapidly and typically converges within approximately 5 iterations, which shows the good convergence property of the optimization algorithm.

Conclusion

This paper proposes a novel PLL method termed SAFE. Unlike previous label-centric approaches that focus on disambiguating noisy candidate labels, SAFE adopts a novel feature-centric perspective, assuming that the candidate labels are correct while the observed features are incomplete. To ensure high-quality feature recovery and trustworthy classifier construction, SAFE explores the intrinsic topological structures within both the feature and label spaces via adaptive local graph learning, and mutually embeds local structures across two spaces. Experiments on some popular datasets demonstrate the effectiveness and superiority of SAFE compared to the state-of-the-art PLL methods, validating the utility of the feature-centric learning paradigm.

Acknowledgments

This work was partially supported by the National Natural Science Foundation of China under Grants Nos. 62576161, 62176116, and 62276136.

References

Chen, B.; Wu, B.; Zareian, A.; Zhang, H.; and Chang, S.-F. 2020. General partial label learning via dual bipartite graph autoencoder. In *AAAI*, volume 34, 10502–10509.

Cour, T.; Sapp, B.; Jordan, C.; and Taskar, B. 2009. Learning from Ambiguously Labeled Images. In *CVPR*, 919–926.

Feng, L.; and An, B. 2018. Leveraging Latent Label Distributions for Partial Label Learning. In *IJCAI*, 2107–2113.

Feng, L.; and An, B. 2019. Partial Label Learning with Self-Guided Retraining. In *AAAI*, volume 33, 3542–3549.

Gong, C.; Liu, T.; Tang, Y.; Yang, J.; Yang, J.; and Tao, D. 2018. A Regularization Approach for Instance-Based Superset Label Learning. *IEEE TCYB*, 48: 967–978.

Gong, X.; Yuan, D.; and Bao, W. 2022. Partial Label Learning via Label Influence Function. In *ICML*, volume 162, 7665–7678.

Guillaumin, M.; Verbeek, J.; and Schmid, C. 2010. Multiple Instance Metric Learning from Automatically Labeled Bags of Faces. In *ECCV*, 634–647.

He, S.; Feng, L.; and Li, L. 2018. Estimating latent relative labeling importances for multi-label learning. In *IEEE ICDM*, 1013–1018.

He, S.; Feng, L.; Lv, F.; Li, W.; and Yang, G. 2022. Partial Label Learning with Semantic Label Representations. In *KDD*, 545–553.

Hou, P.; Geng, X.; and Zhang, M.-L. 2016. Multi-label manifold learning. In *AAAI*, volume 30, 1680–1686.

Huiskes, M. J.; and Lew, M. S. 2008. The MIR flickr Retrieval Evaluation. In *KDD*, 39–43.

Hüllermeier, E.; and Beringer, J. 2006. Learning from Ambiguously Labeled Examples. *IDA*, 10: 419–439.

Jia, X.; Li, Z.; Zheng, X.; Li, W.; and Huang, S.-J. 2019. Label distribution learning with label correlations on local samples. *IEEE TKDE*, 33(4): 1619–1631.

Jia, Y.; Si, C.; and Zhang, M.-L. 2023. Complementary Classifier Induced Partial Label Learning. In *KDD*, 974–983.

Jia, Y.; Yang, F.; and Dong, Y. 2023. Partial Label Learning with Dissimilarity Propagation guided Candidate Label Shrinkage. In *NeurIPS*, volume 36, 34190–34200.

Jiang, B.; Zhang, C.; Wang, Z.; Liang, X.; Zhou, P.; Du, L.; Zhang, Q.; Ding, W.; and Liu, Y. 2025. Scalable fuzzy clustering with collaborative structure learning and preservation. *IEEE TFS*, 33(9): 3047–3060.

Kou, Z.; Qin, S.; Wang, H.; Xie, M.; Chen, S.; Jia, Y.; Liu, T.; Sugiyama, M.; and Geng, X. 2025a. Label Distribution Learning with Biased Annotations by Learning Multi-Label Representation. In *IJCAI*, 2107–2113.

Kou, Z.; Wang, J.; Jia, Y.; and Geng, X. 2025b. Progressive label enhancement. *Pattern Recognition*, 160: 111172.

Liu, L.; and Dietterich, T. G. 2012. A Conditional Multinomial Mixture Model for Superset Label Learning. In *NeurIPS*, volume 25, 557–565.

Liu, L.-P.; and Dietterich, T. G. 2014. Learnability of the Superset Label Learning Problem. In *ICML*, volume 32, 1629–1637.

Lu, Y.; Li, W.; Liu, D.; Li, H.; and Jia, X. 2025. Adaptive-Grained Label Distribution Learning. In *AAAI*, volume 39, 19161–19169.

Lv, J.; Xu, M.; Feng, L.; Niu, G.; Geng, X.; and Sugiyama, M. 2020. Progressive Identification of True Labels for Partial-Label Learning. In *ICML*, volume 119, 6500–6510.

Lyu, G.; Feng, S.; Wang, T.; Lang, C.; and Li, Y. 2021. GM-PLL: Graph Matching Based Partial Label Learning. *IEEE TKDE*, 33: 521–535.

Lyu, G.; Wu, Y.; and Feng, S. 2022. Deep Graph Matching for Partial Label Learning. In *IJCAI*, 3306–3312.

Panis, G.; Lanitis, A.; Tsapatsoulis, N.; and Cootes, T. F. 2016. Overview of Research on Facial Ageing Using the FG-NET Ageing Database. *IET Biom.*, 5: 37–46.

Sheng, M.; Sun, Z.; Cai, Z.; Chen, T.; Zhou, Y.; and Yao, Y. 2024. Adaptive integration of partial label learning and negative learning for enhanced noisy label learning. In *AAAI*, volume 38, 4820–4828.

Si, C.; Jiang, Z.; Wang, X.; Wang, Y.; Yang, X.; and Shen, W. 2024. Partial label learning with a partner. In *AAAI*, volume 38, 15029–15037.

Sun, B.; Deng, Y.; Lin, Y.; Hai, Q.; Yang, Z.; and Lyu, G. 2025. Graph Consistency and Diversity Measurement for Federated Multi-View Clustering. In *AAAI*, volume 39, 20663–20671.

Tang, C.-Z.; and Zhang, M.-L. 2017. Confidence-rated discriminative partial label learning. In *AAAI*, volume 31.

Tian, Y.; Yu, X.; and Fu, S. 2023. Partial label learning: Taxonomy, analysis and outlook. *Neural Networks*, 161: 708–734.

Wang, D.; Zhang, M.; and Li, L. 2022. Adaptive Graph Guided Disambiguation for Partial Label Learning. *IEEE TPAMI*, 44: 8796–8811.

Wang, W.; and Zhang, M. 2022. Partial Label Learning with Discrimination Augmentation. In *KDD*, 1920–1928.

Wang, Z.; and Xu, Y. 2024. A Two-stage Multi-view Partial Multi-label Learning for Enhanced Disambiguation. *KBS*, 293: 111680.

Wu, D.-D.; Wang, D.-B.; and Zhang, M.-L. 2022. Revisiting Consistency Regularization for Deep Partial Label Learning. In *ICML*, volume 162, 24212–24225.

Xu, N.; Lv, J.; and Geng, X. 2019. Partial label learning via label enhancement. In *AAAI*, volume 33, 5557–5564.

Xu, N.; Qiao, C.; Geng, X.; and Zhang, M.-L. 2021. Instance-Dependent Partial Label Learning. In *NeurIPS*, volume 34, 27119–27130.

Yan, Y.; and Guo, Y. 2020. Partial label learning with batch label correction. In *AAAI*, volume 34, 6575–6582.

Yang, F.; Jia, Y.; Liu, H.; Dong, Y.; and Hou, J. 2024. Noisy Label Removal for Partial Multi-Label Learning. In *KDD*, 3724–3735.

Zeng, Z.; Xiao, S.; Jia, K.; Chan, T.; Gao, S.; Xu, D.; and Ma, Y. 2013. Learning by Associating Ambiguously Labeled Images. In *CVPR*, 708–715.

Zhang, C.; Fang, Y.; Liang, X.; Zhang, H.; Zhou, P.; Wu, X.; Yang, J.; Jiang, B.; and Sheng, W. 2024. Efficient Multi view Unsupervised Feature Selection with Adaptive Structure Learning and Inference. In *IJCAI*, 5443–5452.

Zhang, C.; Li, H.; Chen, C.; Qian, Y.; and Zhou, X. 2020. Enhanced group sparse regularized nonconvex regression for face recognition. *IEEE TPAMI*, 44(5): 2438–2452.

Zhang, C.; Li, H.; Gao, Y.; and Chen, C. 2022. Weakly-supervised enhanced semantic-aware hashing for cross-modal retrieval. *IEEE TKDE*, 35(6): 6475–6488.

Zhang, C.; Wang, Z.; Jia, X.; Li, Z.; Chen, C.; and Li, H. 2025a. Multi-view Clustering with Incremental Instances and Views. *IEEE TIP*.

Zhang, C.; Xu, D.; Chen, C.; Zhang, M.; and Li, H. 2025b. Multi-relational multi-view clustering and its applications in cancer subtype identification. *Information Fusion*, 117: 102831.

Zhang, M.; Zhou, B.; and Liu, X. 2016. Partial Label Learning via Feature-Aware Disambiguation. In *KDD*, 1335–1344.

Zhang, M.-L.; and Yu, F. 2015. Solving the Partial Label Learning Problem: An Instance-Based Approach. In *IJCAI*, 4048–4054.

Zhang, M.-L.; Yu, F.; and Tang, C.-Z. 2017. Disambiguation-Free Partial Label Learning. *IEEE TKDE*, 29: 2155–2167.

Zhou, Y.; and Gu, H. 2018. Geometric mean metric learning for partial label data. *IJON*, 275: 394–402.