Classifier-induced Reciprocal Points for Multi-label Open-set Recognition

Yi-Bo Wang[1,2], **Yong Rui**[3], **Min-Ling Zhang**[1,2*]

[1]School of Computer Science and Engineering, Southeast University, Nanjing, China
[2] Key Laboratory of Computer Network and Information Integration, Ministry of Education, China
[3]Lenovo Research, Lenovo Group Ltd., Beijing, China
wang_yb@seu.edu.cn, yongrui@lenovo.com, zhangml@seu.edu.cn

Abstract

Multi-label learning is a practical machine learning paradigm dealing with instances associated with multiple labels simultaneously. Most existing multi-label learning studies are designed under the closed-world assumption, i.e. a fixed size of label space. However, it encounters significant difficulties in open-set scenarios, where test data may contain unknown labels absent from the training set to be recognized. Existing method typically tackles this challenging problem through sub-labeling approximations and prototype-based comparisons, which overlooks the implicit information carried by unknown labels. To address this, we propose a novel framework CREM, i.e. Classifier-induced REciprocal point for Multi-label open-set recognition, which rethinks the above problem from the reciprocal point perspective. Specifically, reciprocal points are formulated by explicitly constraining the opposition feature space to a learnable bounded margin. Then reciprocal points can be induced through the classifier, with the instance-dependent bias eliminated. Subsequently, a unified optimization framework is introduced to jointly facilitate the classifier and reciprocal points induction. Extensive experiments demonstrate the effectiveness and superiority of the proposed CREM approach in the multi-label open-set recognition paradigm.

Introduction

In contrast to traditional multi-class classification (Jia et al. 2023; Chen, Mao, and Zhang 2025), multi-label learning (MLL) (Zhang and Zhou 2013; Liu et al. 2021; Hang and Zhang 2024b) addresses the challenge of learning from multi-semantic objects, where each instance can be associated with multiple relevant labels simultaneously. As multi-semantics objects widely exist in real-world scenarios, such framework has attracted significant attention from diverse communities, such as webpage categorization (Tang et al. 2020) where a new webpage can span several topics, image annotation (You et al. 2020) where an image can represent several scenes, recommendation systems (McAuley, Pandey, and Leskovec 2015) where several personalized items can be recommended concurrently, etc.

Most existing MLL approaches rely on the closed-world assumption (Gao, Xu, and Zhang 2023; Mao, Wang, and Zhang 2023; Shi, Wei, and Li 2024; Gu, Jia, and Zhang 2026), which presumes a fixed label space during both training and testing. However, such an assumption often fails in dynamic environments. In open-world scenarios, particularly where human annotators are not involved during testing, the test set inevitably contains outlier instances associated with previously unknown labels that were not observed during training (Wei, Shi, and Li 2021). For instance, in recommendation systems, new and previously unseen items are continuously introduced, increasing the likelihood of user interactions with items unseen from the training data.

To address this challenge, a more realistic learning paradigm called multi-label open-set recognition (MLOSR) (Wang, Hang, and Zhang 2024) has emerged. MLOSR aims to induce a multi-label prediction model that can correctly classify known labels and effectively recognize unknown labels during test. The main challenges of MLOSR are twofold: (1) the frequent co-occurrence of known and unknown labels significantly complicates the recognition of unknown labels; (2) the label correlations between known and unknown labels may interfere with the accurate classification of known labels. To cope with these problems, the only existing work SLAN recognizes unknown labels by differentiating the sub-labeling information from holistic supervision. It introduces a unified optimization framework to facilitate the known label classification and unknown label recognition. Although SLAN leverages multiple sub-labeling information to approximate potential label information of instances with unknown labels, its fundamental mechanism relies on interpreting holistic supervision as prototype representations derived from the training set (Yang et al. 2020; Liu et al. 2023), performing unknown label recognition through prototype-based comparison.

Such a differentiation strategy may lead to suboptimal performance, as it focuses solely on the known labels, but entirely overlooks the auxiliary information embedded in the unknown labels. Thus, instead of following the above mechanism, we reformulate the problem of MLOSR from the reciprocal point perspective (Chen et al. 2020, 2021), which emphasizes the representation of instances with unknown labels. Specifically, for instances with known label l_k, the majority of unknown instances naturally reside in the non-positive space. Consequently, their features are expected to be more similar to the representations of non-positive in-

*Corresponding author

stances and vice versa. This implies that the latent information associated with unknown labels is implicitly encoded within opposition feature space.

Building upon this perspective, we propose a novel approach named CREM, i.e. ***C**lassifier-induced **RE**ciprocal point for **M**ulti-label open-set recognition*. CREM leverages reciprocal points to explicitly model the separation in the feature space between instances with known and unknown labels. Specifically, CREM first formulates reciprocal points for both the set of positive and negative instances by measuring the distance between the feature representations and the corresponding reciprocal points. Then, the opposition feature space for each positive/negative set is constrained to a learnable bounded margin. Subsequently, the reciprocal points can be induced from a simple linear classifier, with the instance-dependent bias eliminated via an RBF kernel function. Finally, a unified optimization framework is presented to jointly facilitate the classifier and reciprocal points induction. Extensive experiments demonstrate the effectiveness and superiority of the proposed CREM approach under the MLOSR setting.

The rest of this paper is organized as follows. Section 2 briefly reviews related work. Section 3 presents details of the CREM approach. Section 4 reports experimental results. Section 5 concludes this paper.

Related Work

Open-Set Recognition

Inspired by classifiers with rejection option (Da, Yu, and Zhou 2014; Yuan and Wegkamp 2010), open-set recognition (OSR) is formally defined for the first time (Scheirer et al. 2013), with a binary SVM-based framework proposed for its training and evaluation. In recent years, OSR has attracted extensive research attention. Existing approaches can be broadly classified into two categories: discriminative models and generative models.

Among discriminative approaches, (Zhou, Ye, and Zhan 2021) introduces the placeholders-based representative points to mitigate overconfidence in predictions for unknown classes. (Xu, Shen, and Zhao 2023) utilizes supervised contrastive learning to improve the quality of learned representations. To address the limitations of traditional OSR metrics, OpenAUC (Wang et al. 2022) serves as a novel metric tailored for open-set scenarios. (Wang et al. 2024) encourages diverse representations, effectively reducing open space risk, via expert-specific attention regularization and adaptive logit fusion. (Hang and Zhang 2024a) tackles class imbalance and representation compromise in open-set semi-supervised learning, via adaptive logit adjustment and label-specific feature learning, respectively.

In contrast, generative approaches model the decision boundary between known and unknown classes by synthesizing instances using various techniques, including GANs, autoencoders, prototypes, and reciprocal points. (Neal et al. 2018) utilizes GANs (Goodfellow et al. 2014) to generate instances similar to training data but not belonging to any specific classes, effectively enriching the training dataset. (Oza and Patel 2019) proposes a two-stage approach, where an encoder is first trained for closed-set recognition, followed by a decoder incorporating class-conditional information for unknown classes detection. (Sun et al. 2020) employs a variational autoencoder (VAE) to align latent features with different Gaussian priors, facilitating unknown classes detection. (Guo et al. 2021) combines VAE and capsule network to enforce intra-class feature consistency through alignment with a predefined distribution. GCPL (Yang et al. 2020), a prototype-based classification method tailored for open-world scenarios, replaces the Softmax classifier. (Liu et al. 2023) introduces a label-to-prototype mapping function to construct prototypes for both known classes and unknown classes. Furthermore, RPL (Chen et al. 2020) is developed to identify unknown classes based on the deviation from reciprocal points. Building upon this, (Chen et al. 2021) incorporates adversarial enhancement confusing training instances via confrontation between known data and reciprocal points, thereby enhancing the model's discriminative capacity.

Multi-Label Open Set Recognition

In the past decades, a wide range of approaches have been proposed to tackle multi-label learning problem (Zhang and Zhou 2013; Liu et al. 2021). To address the challenge of an exponential-sized output space, modeling label correlations has become a mainstream strategy. Roughly speaking, extant approaches can be grouped into three categories based on the order of correlations modeling, i.e. *first-order* approaches, *second-order* approaches and *high-order* approaches. First-order approaches tackle multi-label learning problem by treating each label independently (Boutell et al. 2004; Zhang et al. 2018; Mao, Hang, and Zhang 2024). Second-order approaches focus on pairwise interactions modeling among class labels (Fürnkranz et al. 2008; Brinker, Mencía, and Fürnkranz 2014). High-order approaches capture relationships among a subset of or all class labels (Huang et al. 2019; Kou et al. 2024; Mao, Rui, and Zhang 2025).

To bridge the gap between multi-label learning and open-world scenarios, multi-label open-set recognition has emerged in recent years. To the best of our knowledge, SLAN (Wang et al. 2024) corresponds to the only prior work. Specifically, SLAN enhances the sub-labeling information by leveraging the feature structural information, and recognizes unknown labels by differentiating the sub-labeling information from holistic supervision. However, the above approach incorporates multiple sub-labeling information to approximate potential label information might be suboptimal as it only focuses on the known labels, but entirely overlooks the auxiliary information embedded in the unknown labels. In this paper, we introduce the CREM approach, the first attempt to tackle MLOSR from the perspective of reciprocal points, which emphasizes the representation of instances with unknown labels. The proposed CREM approach will be introduced in the next section.

The CREM Approach

Preliminaries

Let $\mathcal{X} = \mathbb{R}^d$ denote the d-dimensional feature space and $\mathcal{Y} = \{l_1, l_2, \ldots, l_q\}$ denote the label space with q class la-

bels. Each multi-label instance is represented as $(\boldsymbol{x}_i, Y_i)$, where $\boldsymbol{x}_i \in \mathcal{X}$ is a real-valued feature vector and $Y_i \subseteq \mathcal{Y}$ is the set of relevant labels associated with $\boldsymbol{x}_i$. For notation simplicity, a q-dimensional indicator vector $\boldsymbol{y}_i = [y_{i1}, y_{i2}, \ldots, y_{iq}]^\top \in \{-1, 1\}^q$ is utilized to instantiate the set of relevant labels Y_i, where $y_{ik} = 1$ indicates $l_k \in Y_i$ and $y_{ik} = -1$ otherwise. By arranging feature vectors and label vectors of m training instances, we obtain the feature matrix $\mathbf{X} = [\boldsymbol{x}_1, \ldots, \boldsymbol{x}_m]$ and label matrix $\mathbf{Y} = [\boldsymbol{y}_1, \ldots, \boldsymbol{y}_m]^\top$.

Given a multi-label training set $\mathcal{D} = \{(\boldsymbol{x}_i, Y_i) \mid 1 \leq i \leq m\}$, the goal of MLOSR is to learn a measurable multi-label recognition function $f : \mathcal{X} \to 2^{\mathcal{Y}}$ from $\mathcal{D}$ that minimizes the following **Open Set Risk**:

$$\arg\min_{f \in \mathcal{H}} R_\varepsilon(f(\mathcal{D})) + \lambda R_\mathcal{O}(f), \tag{1}$$

where R_ε measures the empirical risk on instances with known labels, and $R_\mathcal{O}$ measures the open space risk of instances with potential unknown labels on open space $\mathcal{O}$. That is, solving the MLOSR problem in Eq.(1) is equivalent to correctly classifying known labels for each instance and recognizing unknown labels within its relevant label set.

Reciprocal Points for Open space risk Management

With respect to the k-th label l_k, the set of positive training instances $\mathcal{P}_k$ as well as the set of negative training instances $\mathcal{N}_k$ without considering unknown labels are determined based on the relevance between the instances and the label l_k:

$$\begin{aligned} \mathcal{P}_k &= \{\boldsymbol{x}_i \mid (\boldsymbol{x}_i, Y_i) \in \mathcal{D}, l_k \in Y_i\}, \\ \mathcal{N}_k &= \{\boldsymbol{x}_i \mid (\boldsymbol{x}_i, Y_i) \in \mathcal{D}, l_k \notin Y_i\}. \end{aligned} \tag{2}$$

The **reciprocal point** $\boldsymbol{p}_k$ characterizes the instantiated representation of instances of the latent unknown label space. Thus, instances that do not belong to $\mathcal{P}_k$ should be closer to this reciprocal point, which can be formulated as:

$$\forall \boldsymbol{x} \in \mathcal{P}_k, \boldsymbol{z} \notin \mathcal{P}_k, \mathbf{d}^2(\boldsymbol{z}, \boldsymbol{p}_k) \leq \mathbf{d}^2(\boldsymbol{x}, \boldsymbol{p}_k), \tag{3}$$

where $\mathbf{d}(\cdot, \cdot)$ returns the distance between two vectors and is set to the Euclidean metric in this paper. Here, $\boldsymbol{z}$ represents the instance that belongs to $\mathcal{N}_k$ or the instance associated with unknown labels.

Intuitively, limited open space allows the management of open space risk. By constraining the maximum distance between the instance $\boldsymbol{z}$ and its reciprocal point $\boldsymbol{p}_k$, the open space can be effectively limited within a bounded range. However, such a strategy is infeasible during training, as instances associated with potential unknown labels are unseen. Fortunately, since the space instantiated by $\mathcal{P}_k$ and the open space are complementary to each other, the open space can be indirectly bounded by constraining the distance between the instances from $\mathcal{P}_k$ and the corresponding reciprocal point $\boldsymbol{p}_k$ as follows:

$$\min_{\boldsymbol{p}_k, R_k} \frac{1}{m} \sum_{i=1}^{m} \mathbb{I}(y_{ik} = 1) \ell(R_k^2 - \mathbf{d}^2(\boldsymbol{x}, \boldsymbol{p}_k)), \tag{4}$$

where $\ell(\cdot)$ is a surrogate loss function and R_k is a learnable margin. Specifically, Eq.(4) is equivalent to making the distance between $\boldsymbol{z}$ and $\boldsymbol{p}_k$ smaller than R_k.

Reciprocal Points Induction

Similar to prototype learning, the reciprocal point $\boldsymbol{p}_k$ effectively captures the opposite characteristics of the positive label, providing appropriate distinguishing information to facilitate its discrimination, and the same holds for $\boldsymbol{n}_k$ and $\mathcal{N}_k$. Thus, it is necessary to analyze the underlying properties of the distance relation derived from $\boldsymbol{p}_k$. Then, we can rewrite Eq.(3) as:

$$-\boldsymbol{p}_k^\top \boldsymbol{x} + \frac{1}{2}\boldsymbol{x}^\top \boldsymbol{x} \geq -\boldsymbol{p}_k^\top \boldsymbol{z} + \frac{1}{2}\boldsymbol{z}^\top \boldsymbol{z}, \tag{5}$$

which indicates the distance relation is governed by the inner product between the reciprocal point and the instance, along with an instance-dependent bias term.

Let $\phi(\cdot)$ be a nonlinear mapping implemented by the RBF kernel function κ and $\phi(\cdot) \in \mathbb{R}^{\mathcal{H}_\kappa}$. Instance-dependent bias can be eliminated by mapping the original input space to the kernel space since $\kappa(\boldsymbol{x}, \boldsymbol{x}) = \phi(\boldsymbol{x})^\top \phi(\boldsymbol{x}) = 1$. Then, Eq.(5) can be simplified through a scaling transformation as:

$$\boldsymbol{w}_k^\top \phi(\boldsymbol{x}) + b_k \geq 1, \quad \boldsymbol{w}_k^\top \phi(\boldsymbol{z}) + b_k \leq -1, \tag{6}$$

where $\boldsymbol{w}_k = -\boldsymbol{p}'_k \in \mathbb{R}^{\mathcal{H}_\kappa}$. This implies that the reciprocal point $\boldsymbol{p}'_k$ can be induced via the classifier determined by $\boldsymbol{w}_k$ and b_k. Subsequently, by the property of ℓ_2-norm, we have

$$\begin{aligned} \|\boldsymbol{w}_k - \phi(\boldsymbol{x})\|_2 &= \|\boldsymbol{w}_k + \phi(\boldsymbol{x}) - 2\phi(\boldsymbol{x})\|_2 \\ &\leq \|\boldsymbol{p}'_k - \phi(\boldsymbol{x})\|_2 + 2\|\phi(\boldsymbol{x})\|_2 \\ &\leq R_k + 2. \end{aligned} \tag{7}$$

That is, given the induced $\boldsymbol{p}'_k$, the reciprocal point $\boldsymbol{n}'_k$, defined by $\boldsymbol{n}'_k = \boldsymbol{w}_k$, is directly induced to manage the open space in Eq.(4) w.r.t $\mathcal{N}_k$, thereby significantly reducing computational complexity.

Here, we assign a simple ridge regression model to each label space as the classifier over the training set $\mathcal{D}$:

$$\min_{\mathbf{W}, \boldsymbol{b}} \sum_{i=1}^{m} \|\mathbf{W}^\top \phi(\boldsymbol{x}_i) + \boldsymbol{b} - \boldsymbol{y}_i\|_2^2 + \lambda_1 \|\mathbf{W}\|_\mathrm{F}^2. \tag{8}$$

Here, $\mathbf{W} = [\boldsymbol{w}_1, \ldots, \boldsymbol{w}_q] \in \mathbb{R}^{\mathcal{H}_\kappa \times q}$ is the predictive modeling coefficients and $\boldsymbol{b} = [b_1, b_2, \ldots, b_q]^\top \in \mathbb{R}^q$ is the intercept to be determined. λ_1 is a trade-off parameter for regularization. However, the above predictive model actually deals with the q labels independently. To exploit the intrinsic label correlations among multi-label instances, we incorporate a label correlation matrix $\mathbf{C} = [c_{ij}]_{q \times q}$ to represent the similarity of between pairs of labels:

$$\min_{\mathbf{W}} \sum_{i=1}^{q} \sum_{j=1}^{q} c_{ij} \|\boldsymbol{w}_i - \boldsymbol{w}_j\|_2^2. \tag{9}$$

Each element in $\mathbf{C}$ is calculated as: $c_{ij} = \frac{1}{2}[P(l_j \mid l_i) + P(l_i \mid l_j)]$, where $P(l_j \mid l_i)$ is the probability that label l_j appears when label l_i appears and the diagonal elements of similarity matrix $\mathbf{C}$ are set to 0. We calculate the similarity matrix $\mathbf{C}$ from the training set. Then, incorporating the Eq.(4), Eq.(8) and Eq.(9), the overall CREM framework can

be achieved as follows:

$$\min_{\mathbf{W},\boldsymbol{b},\mathbf{P},\boldsymbol{R}} \frac{1}{2}\sum_{i=1}^{m}\|\mathbf{W}^\top\phi(\boldsymbol{x}_i)+\boldsymbol{b}-\boldsymbol{y}_i\|_2^2+\frac{\lambda_1}{2}\|\mathbf{W}\|_{\mathrm{F}}^2 +\frac{\lambda_2}{2}tr(\mathbf{W}\mathbf{L}\mathbf{W}^\top)+\frac{\lambda_3}{2}\|\mathbf{W}-\mathbf{P}\|_{\mathrm{F}}^2 +\frac{\alpha}{2m}\sum_{k=1}^{q}\sum_{i=1}^{m}\mathbb{I}(y_{ik}=1)\ell(R_k^2-\mathbf{d}^2\left(\phi(\boldsymbol{x}_i),\boldsymbol{p}_k'\right)), \tag{10}$$

where $\mathbf{P}=[-\boldsymbol{p}_1',\ldots,-\boldsymbol{p}_q']\in\mathbb{R}^{\mathcal{H}_\kappa\times q}$ is the reciprocal points matrix and $\boldsymbol{R}=[R_1,\ldots,R_q]^\top\in\mathbb{R}^q$ is the learnable margin to be determined. $\mathbf{L}=diag(\mathbf{C}\mathbf{1}_q)-\mathbf{C}$ denotes the Laplacian matrix and $\ell(\cdot)$ corresponds to the hinge loss function. Here, the first three terms control the empirical risk, the fourth term penalizes the difference between $\mathbf{W}$ and $\mathbf{P}$, and the last term controls the open space risk.

It is worth noting that, in the optimization problem Eq.(10), $\phi(\cdot)$ is the nonlinear mapping implemented by the RBF kernel function κ. Therefore, we cannot obtain an explicit solution of $\mathbf{W}$. According to the Representer Theorem (Schölkopf and Smola 2002), the predictive model can be expressed as a linear combination of the training instances. Let $\mathbf{\Phi}=[\phi(\boldsymbol{x}_1),\ldots,\phi(\boldsymbol{x}_m)]^\top\in\mathbb{R}^{m\times\mathcal{H}_\kappa}$ denote the nonlinear mapping instance matrix. We can then express $\boldsymbol{w}_k$ as $\boldsymbol{w}_k=\sum_{i=1}^{m}\theta_{ki}\phi(\boldsymbol{x}_i)=\mathbf{\Phi}^\top\boldsymbol{\theta}_{\mathbf{W}}^k$ and then $\mathbf{W}=\mathbf{\Phi}^\top\mathbf{\Theta}_{\mathbf{W}}$, where $\mathbf{\Theta}_{\mathbf{W}}=\left[\boldsymbol{\theta}_{\mathbf{W}}^1,\ldots,\boldsymbol{\theta}_{\mathbf{W}}^q\right]\in\mathbb{R}^{m\times q}$ represents the combination coefficients to be determined. Meanwhile, since reciprocal points are induced by the classifier, $\mathbf{P}$ would have similar representation to $\mathbf{W}$, that is, $\mathbf{P}=\mathbf{\Phi}^\top\mathbf{\Theta}_{\mathbf{P}}$, where $\mathbf{\Theta}_{\mathbf{P}}=\left[\boldsymbol{\theta}_{\mathbf{P}}^1,\ldots,\boldsymbol{\theta}_{\mathbf{P}}^q\right]\in\mathbb{R}^{m\times q}$.

By substituting $\mathbf{W}=\mathbf{\Phi}^\top\mathbf{\Theta}_{\mathbf{W}}$ and $\mathbf{P}=\mathbf{\Phi}^\top\mathbf{\Theta}_{\mathbf{P}}$ into the objective function Eq.(10), the overall CREM framework can be achieved as follows:

$$\min_{\substack{\mathbf{\Theta}_{\mathbf{W}},\boldsymbol{b},\\ \mathbf{\Theta}_{\mathbf{P}},\boldsymbol{R}}} \frac{1}{2}\|\mathbf{K}\mathbf{\Theta}_{\mathbf{W}}+\mathbf{1}_m\boldsymbol{b}^\top-\mathbf{Y}\|_{\mathrm{F}}^2+\frac{\lambda_1}{2}\|\mathbf{\Phi}^\top\mathbf{\Theta}_{\mathbf{W}}\|_{\mathrm{F}}^2 +\frac{\lambda_2}{2}tr(\mathbf{\Phi}^\top\mathbf{\Theta}_{\mathbf{W}}\mathbf{L}\mathbf{\Theta}_{\mathbf{W}}^\top\mathbf{\Phi})+\frac{\lambda_3}{2}\|\mathbf{\Phi}^\top(\mathbf{\Theta}_{\mathbf{W}}-\mathbf{\Theta}_{\mathbf{P}})\|_{\mathrm{F}}^2 +\frac{\alpha}{2m}\sum_{k=1}^{q}\sum_{i=1}^{m}\mathbb{I}(y_{ik}=1)\ell(R_k^2-\mathbf{d}^2\left(\phi(\boldsymbol{x}_i),-\mathbf{\Phi}^\top\boldsymbol{\theta}_{\mathbf{P}}^k\right)). \tag{11}$$

Unknown Labels Recognition

As the classifier is exclusively trained on the data with known label, it tends to assigns relatively high positive/negative label confidence to the training data. However, latent correlations between known and unknown labels might mislead the classifier, resulting in relatively lower label confidence. According to Eq.(3), label confidence is related to the maximum distance between an instance and its reciprocal points, i.e. $max(\mathbf{d}(\boldsymbol{p}_k,\cdot),\mathbf{d}(\boldsymbol{n}_k,\cdot))$. Consequently, the unseen instance $\boldsymbol{u}$ associated with unknown labels would be closer to all reciprocal points, resulting in lower confidence scores. To quantify label confidence, we employ q logistic functions to derive label confidence scores $\mathcal{S}_k(\boldsymbol{u})\in[0,1]$, where each function maps the maximum distance metric to a calibrated value. Assuming that high confidence tends to be concentrated on a subset of labels, we aggregate the top-K maximum $\mathcal{S}_k(\boldsymbol{u})$ to obtain the final result. The threshold τ is chosen so that a high fraction (e.g. 95%) of the training data is accepted. Instances with scores lower than τ are recognized to be associated with unknown labels.

$$\mathcal{S}(\boldsymbol{u})=\frac{1}{K}\sum_{k}\mathrm{TopK}(\mathcal{S}_k(\boldsymbol{u})) \tag{12}$$

Optimization

Obviously, the interaction between variables prevents them from being calculated simultaneously. Consequently, in this paper, we iteratively optimize one set of parameters while keeping the other set of parameters fixed until convergence.

Update $\mathbf{\Theta}_{\mathbf{W}}$ and $\boldsymbol{b}$. While $\mathbf{\Theta}_{\mathbf{P}}$ and $\boldsymbol{R}$ are fixed, the optimization problem in Eq.(11) can be stated as follows:

$$\min_{\mathbf{\Theta}_{\mathbf{W}},\boldsymbol{b}} \frac{1}{2}\|\mathbf{K}\mathbf{\Theta}_{\mathbf{W}}+\mathbf{1}_m\boldsymbol{b}^\top-\mathbf{Y}\|_{\mathrm{F}}^2+\frac{\lambda_1}{2}\|\mathbf{\Phi}^\top\mathbf{\Theta}_{\mathbf{W}}\|_{\mathrm{F}}^2 +\frac{\lambda_2}{2}tr(\mathbf{\Phi}^\top\mathbf{\Theta}_{\mathbf{W}}\mathbf{L}\mathbf{\Theta}_{\mathbf{W}}^\top\mathbf{\Phi})+\frac{\lambda_3}{2}\|\mathbf{\Phi}^\top(\mathbf{\Theta}_{\mathbf{W}}-\mathbf{\Theta}_{\mathbf{P}})\|_{\mathrm{F}}^2. \tag{13}$$

Let $\mathbf{K}=\mathbf{\Phi}\mathbf{\Phi}^\top\in\mathbb{R}^{m\times m}$ represents the kernel matrix with (i,j)-th element $\mathbf{K}_{ij}=\kappa(\boldsymbol{x}_i,\boldsymbol{x}_j)$. Then the first-order derivative of Eq.(13) w.r.t $\mathbf{\Theta}_{\mathbf{W}}$ is

$$\begin{aligned}\nabla\mathbf{\Theta}_{\mathbf{W}}=&\mathbf{K}^\top(\mathbf{K}\mathbf{\Theta}_{\mathbf{W}}+\mathbf{1}_m\boldsymbol{b}^\top-\mathbf{Y})+\lambda_1\mathbf{K}\mathbf{\Theta}_{\mathbf{W}}\\&+\lambda_2\mathbf{K}\mathbf{\Theta}_{\mathbf{W}}\mathbf{L}+\lambda_3\mathbf{K}(\mathbf{\Theta}_{\mathbf{W}}-\mathbf{\Theta}_{\mathbf{P}}).\end{aligned} \tag{14}$$

Setting the above Eq.(14) to $\mathbf{0}$, we have

$$(\mathbf{K}+(\lambda_1+\lambda_3)\mathbf{I}_{m\times m})\mathbf{\Theta}_{\mathbf{W}}+\lambda_2\mathbf{\Theta}_{\mathbf{W}}\mathbf{L}=\lambda_1\mathbf{\Theta}_{\mathbf{P}}+\mathbf{Y}-\mathbf{1}_m\boldsymbol{b}^\top, \tag{15}$$

which is a Sylvester equation and can be solved by any off-the-shelf solvers (Wei, Dobigeon, and Tourneret 2015).

Similarly, setting the first-order derivative of Eq.(13) w.r.t $\boldsymbol{b}$ to $\mathbf{0}$, we have the closed-form solution

$$\boldsymbol{b}=-\frac{1}{n}(\mathbf{K}\mathbf{\Theta}_{\mathbf{W}}-\mathbf{Y})^\top\mathbf{1}_m. \tag{16}$$

Update $\mathbf{\Theta}_{\mathbf{P}}$. When $\mathbf{\Theta}_{\mathbf{W}}$, $\boldsymbol{b}$ and $\boldsymbol{R}$ are fixed, since the reciprocal point $\boldsymbol{p}_k$ of each label is independent with others, we can re-write the optimization problem Eq.(11) as follows:

$$\min_{\boldsymbol{\theta}_{\mathbf{P}}^k}\frac{\lambda_3}{2}\|\mathbf{\Phi}^\top(\boldsymbol{\theta}_{\mathbf{W}}^k-\boldsymbol{\theta}_{\mathbf{P}}^k)\|_2^2 +\frac{\alpha}{2m}\sum_{i=1}^{m}\mathbb{I}(y_{ik}=1)\ell(R_k^2-\mathbf{d}^2\left(\phi(\boldsymbol{x}_i),-\mathbf{\Phi}^\top\boldsymbol{\theta}_{\mathbf{P}}^k\right)). \tag{17}$$

Setting the first-order derivative of Eq.(17) w.r.t $\boldsymbol{\theta}_{\mathbf{P}}^k$ to $\mathbf{0}$, we have the closed-form solution

$$\boldsymbol{\theta}_{\mathbf{P}}^k=\frac{1}{Z}(\lambda_3\boldsymbol{\theta}_{\mathbf{W}}^k+\frac{\alpha}{m}\sum_{j=1}^{m}\delta_{kj}\mathbf{I}_j). \tag{18}$$

Compared Algorithms	Datasets						
	enron	recreation	slashdot	arts	education	rcvsubset2-2	bibtex
				AUROC↑			
OC-SVM	0.452±0.080 •	0.497±0.014 •	0.468±0.063 •	0.521±0.036 •	0.520±0.024 •	0.508±0.016 •	0.542±0.026 •
IFOREST	0.462±0.075 •	0.496±0.014 •	0.469±0.064 •	0.522±0.033 •	0.522±0.027 •	0.509±0.016 •	0.543±0.026 •
MUENL-F	0.470±0.075 •	0.499±0.015 •	0.444±0.060 •	0.523±0.037 •	0.523±0.026 •	0.507±0.015 •	0.533±0.027 •
SLAN	0.542±0.185	**0.604±0.029**	**0.624±0.013** ○	**0.628±0.026**	0.580±0.020	0.524±0.277	0.529±0.014 •
CREM	**0.592±0.023**	0.591±0.030	0.562±0.029	**0.628±0.018**	**0.599±0.017**	**0.609±0.022**	**0.567±0.024**
				AUPR↑			
OC-SVM	0.794±0.049	0.848±0.030 •	0.771±0.048	0.868±0.041	0.856±0.058	0.756±0.008 •	0.629±0.027
IFOREST	0.797±0.050	0.847±0.030 •	0.771±0.045	0.869±0.042	0.856±0.058	0.755±0.010 •	0.631±0.027
MUENL-F	0.788±0.051	0.844±0.030 •	0.726±0.052 •	0.863±0.041 •	0.851±0.057 •	0.746±0.008 •	0.603±0.026 •
SLAN	0.825±0.082	**0.883±0.025**	**0.843±0.045**	0.894±0.030	0.874±0.042	0.773±0.125	0.611±0.029 •
CREM	**0.840±0.077**	0.877±0.028	0.806±0.046	**0.898±0.028**	**0.879±0.048**	**0.793±0.013**	**0.648±0.023**

Table 1: Open-set recognition performance of each comparing approaches (mean±std) in terms of *AUROC* and *AUPR*. ↑ (↓) indicates the larger (smaller) the value, the better the performance. The best and second best results are highlighted in **boldface** and underline, respectively. In addition, •/○ indicates whether CREM is statistically superior/inferior to the comparing approaches on each data set with pairwise t-test (at 0.05 significance level).

Here, $\mathbf{I}_j$ corresponds to the j-th column of the identity matrix and $Z = \lambda_1 - \frac{\alpha}{m}\sum_{j=1}^{m}\delta_{kj}$ is the normalization factor with $\delta_{kj} = \mathbb{I}(y_{ik} = 1)\sigma_{ki}$, which is a piecewise function defined as follows:

$$\sigma_{ki} = \begin{cases} -1, & \text{if } R_k^2 - \mathbf{d}^2(-\mathbf{\Phi}^\top \boldsymbol{\theta}_{\mathbf{P}}^k, \phi(\boldsymbol{x}_j)) \leq 1, \\ 0, & otherwise. \end{cases} \tag{19}$$

Algorithm 1: The pseudo-code of CREM

Input:
$\mathcal{D}$: multi-label training set;
$\lambda_1, \lambda_2, \lambda_3, \alpha$: trade-off parameters;
K: the number of maximum $\mathcal{S}_k(\cdot)$;
$\boldsymbol{u}$: unseen instance ($\boldsymbol{u} \in \mathcal{X}$);
Output: The predicted label set $Y(\boldsymbol{u})$ and the recognition score $\mathcal{S}(\boldsymbol{u})$.
Process:
1: Instantiate the $\mathbf{\Theta_W}$, $\mathbf{\Theta_P}$ with $\mathbf{0}_{m\times q}$;
2: Calculate the kernel matrix $\mathbf{K} = [\kappa(\boldsymbol{x}_i, \boldsymbol{x}_j)]_{m\times m}$;
3: Calculate similarity matrix $\mathbf{C}$;
4: **repeat**
5: Update $\mathbf{\Theta_W}$ and $\boldsymbol{b}$ according to Eq.(15) and Eq.(16);
6: Update $\mathbf{\Theta_P}$ according to Eq.(18);
7: Update R_k according to Eq.(21);
8: **until** convergence
9: **return** $\mathcal{S}(\boldsymbol{u})$ according to Eq.(12) and $Y(\boldsymbol{u}) = \{l_k | \sum_{i=1}^{m} \theta_{ki}\kappa(\boldsymbol{x}_i, \boldsymbol{u}) > 0\}$.

Update R. While $\mathbf{\Theta_W}$, $\mathbf{\Theta_P}$ and $\boldsymbol{b}$ are fixed, since R_k is independent with others, the optimization problem Eq.(11) can be stated as follows:

$$\min_{R_k} \frac{\alpha}{2m}\sum_{i=1}^{m}\mathbb{I}(y_{ik} = 1)\ell(R_k^2 - \mathbf{d}^2\left(\phi(\boldsymbol{x}_i), -\mathbf{\Phi}^\top\boldsymbol{\theta}_{\mathbf{P}}^k\right)). \tag{20}$$

The above optimization problem can be solved by updating R_k with gradient descent. Specifically, the gradient of the objective function with respect to R_k is

$$\nabla R_k = \frac{\alpha}{m}\sum_{i=1}^{m}\mathbb{I}(y_{ik} = 1)\sigma_{ki}. \tag{21}$$

The complete procedure of CREM is summarized in Algorithm 1. First, the combination coefficients $\mathbf{\Theta_W}$ and $\mathbf{\Theta_P}$ are initialized (Step 1). Then, the kernel matrix $\mathbf{K}$ is constructed using a predefined kernel function (Step 2) and the similarity matrix $\mathbf{C}$ is computed to exploit label correlations (Step 3). After that, the optimization procedure is performed (Step 4-8). Finally, the relevant label set of the unseen instance is predicted and the recognition score is generated (Step 9).

Dataset	$\lvert\mathcal{S}\rvert$	$dim(\mathcal{S})$	$L(\mathcal{S})$	$LCard(\mathcal{S})$
enron	1702	1001	24	3.124
slashdot	3659	1079	14	1.173
recreation	5000	606	15	1.361
arts	5000	462	14	1.512
education	5000	550	11	1.374
rcvsubset2-2	6000	944	39	2.170
bibtex	7395	1835	27	0.954

Table 2: Characteristics of experimental data sets.

Experiments

Experimental Setup

Datasets. Following the experimental protocol proposed in (Wang, Hang, and Zhang 2024), seven real-world multi-label tabular datasets are employed for comparative studies, whose characteristics such as feature sparsity and feature-engineering nuances differ from vision domains. For each dataset, 50% labels are selected as known labels and the remaining labels as unknown labels. Table 2 summarizes the detailed characteristics of each benchmark multi-label data set $\mathcal{S}$ employed in the experiments, including the number of instances $|\mathcal{S}|$, number of features $dim(\mathcal{S})$, number of class labels $L(\mathcal{S})$, label cardinality $LCard(\mathcal{S})$.

Compared Algorithms	Datasets						
	enron	recreation	slashdot	arts	education	rcvsubset2-2	bibtex
	Macro-averaging AUC↑						
LIFT	0.646±0.022 •	0.635±0.010 •	0.596±0.012 •	0.630±0.014 •	0.623±0.010 •	0.705±0.015 •	0.703±0.027 •
MUENL-P	0.555±0.010 •	0.549±0.009 •	0.524±0.014 •	0.554±0.013 •	0.551±0.014 •	0.554±0.010 •	0.523±0.011 •
SENCE	0.659±0.033 •	0.655±0.015 •	0.653±0.016 •	0.645±0.014 •	0.660±0.012 •	0.744±0.015 •	0.583±0.045 •
LIMIC	0.560±0.014 •	0.538±0.010 •	0.552±0.011 •	0.545±0.015 •	0.549±0.019 •	0.520±0.015 •	0.588±0.015 •
SLAN	0.677±0.030 •	0.670±0.015	0.687±0.020	0.655±0.013	0.674±0.016	0.764±0.014 •	**0.827±0.016**
CREM	**0.704±0.019**	**0.677±0.012**	**0.704±0.018**	**0.659±0.014**	**0.676±0.014**	**0.785±0.014**	0.820±0.016
	Average Precision↑						
LIFT	0.655±0.025	0.653±0.015 •	0.541±0.018 •	0.551±0.035	0.678±0.036	0.473±0.013 •	0.606±0.028 •
MUENL-P	0.590±0.010 •	0.551±0.021 •	0.454±0.013 •	0.488±0.029 •	0.626±0.027 •	0.348±0.012 •	0.352±0.009 •
SENCE	0.663±0.023	0.672±0.021	0.574±0.017 •	0.566±0.036	0.701±0.033	0.513±0.017 •	0.601±0.039 •
LIMIC	0.638±0.020 •	0.615±0.025 •	0.518±0.020 •	0.483±0.051 •	0.638±0.043 •	0.393±0.060 •	0.542±0.022 •
SLAN	0.664±0.025	0.666±0.018 •	**0.614±0.023**	**0.576±0.034**	**0.709±0.033**	0.532±0.017	**0.770±0.030** ◦
CREM	**0.670±0.026**	**0.683±0.016**	0.611±0.018	0.567±0.044	0.705±0.036	**0.540±0.024**	0.690±0.030

Table 3: Open-set recognition performance of each comparing approaches (mean±std) in terms of *Macro-averaging AUC* and *Average Precision*. ↑ (↓) indicates the larger (smaller) the value, the better the performance. The best and second best results are highlighted in **boldface** and underline, respectively. In addition, •/◦ indicates whether CREM is statistically superior/inferior to the comparing approaches on each data set with pairwise t-test (at 0.05 significance level).

Evaluation Metrics. *AUROC* and *AUPR* are employed to evaluate the open-set recognition performance. In addition, to evaluate the performance of multi-label classification, five multi-label evaluation metrics are utilized, including *Macro-averaging AUC*, *Average precision*, *Ranking loss*, *One-error* and *Coverage*. Detailed definitions of these metrics can be found in (Wang et al. 2022; Zhang and Zhou 2013).

Compared Methods. We compare CREM with four open-set recognition methods, including OC-SVM (Ma and Perkins 2003), IFOREST (Liu, Ting, and Zhou 2008), MUENL-F (Zhu, Ting, and Zhou 2018) and SLAN (Wang, Hang, and Zhang 2024), where MUENL-F and SLAN are tailored for MLOSR. Besides, we compare CREM with five state-of-the-art multi-label learning methods, including LIFT (Zhang and Wu 2015), MUENL-P(Zhu, Ting, and Zhou 2018), SENCE (Wang, Hang, and Zhang 2022), LIMIC (Mao, Wang, and Zhang 2023) and SLAN (Wang, Hang, and Zhang 2024). More details about these compared methods can be found in Appendix A.1.

Configuration. For the proposed CREM approach, we employ RBF kernel function and set the parameters as follows: the trade-off parameters $\lambda_1 = 1, \lambda_2 = 0.1, \lambda_3 = 10, \alpha = 1$. K is searched in $\{2, \ldots, q-1\}$. Detailed discussion about the choice of these parameters can be found in the Sensitivity analysis paragraph and Appendix B.

Empirical Results

Table 1 and Table 3 report detailed empirical results in terms of open-set recognition and multi-label learning. Furthermore, pairwise t-test (Demsar 2006) (0.05 significance level) is conducted to analyze whether the proposed CREM performs statistically better than other comparing algorithms. Results of statistical test are reported in Table 4. The results on other multi-label evaluation metrics can be found in Appendix A.2. Based on these results, it is impressive to observe that:

- Across all evaluation metrics, CREM achieves the best or second best performance in 95.9% cases over all the 7 data sets.
- As shown in Table 4, CREM achieves statistically superior or at least comparable performance against comparing open-set recognition approaches in all open-set evaluation metrics. The superior performance of CREM indicates that it is a promising direction to facilitate MLOSR with the strategy of classifier-induced reciprocal points.
- Meanwhile, across all multi-label evaluation metrics, CREM achieves superior or at least comparable performance against the comparing approaches in 97.7% cases. These results demonstrate the effectiveness of our proposed method to correctly classify known labels.

Metrics	CREM **against**			
	OC-SVM	IFOREST	MUENL-F	SLAN
AUROC	7/0/0	7/0/0	7/0/0	1/5/1
AUPR	2/5/0	2/5/0	6/1/0	1/6/0
In Total	**9/5/0**	**9/5/0**	**13/1/0**	**2/11/1**

Table 4: Win/tie/loss counts (pairwise t-test at 0.05 significant level) for CREM against other comparing approaches.

Further Analysis

Ablation study. Ablation studies are conducted to analyze the effectiveness of two essential components: (1) kernel space mapping; (2) classifier-induced reciprocal points. Accordingly, we design two degenerate variants named CREM-C and CREM-K for performance comparison:

- CREM-C employs Eq.(5) instead of Eq.(6) to remove the kernel space mapping procedure in CREM, which corresponds to the degenerate case inducing reciprocal points via classifier in original feature space with the instance-dependent bias term.

Compared Algorithms	Datasets						
	enron	recreation	slashdot	arts	education	rcvsubset2-2	bibtex
				AUROC↑			
CREM-C	0.547±0.070	0.502±0.013 •	0.515±0.040 •	0.483±0.023 •	0.496±0.020 •	0.500±0.011 •	0.458±0.025 •
CREM-K	0.589±0.032	0.584±0.032	0.510±0.009 •	0.619±0.020 •	0.590±0.018 •	0.591±0.023	0.556±0.024
CREM	**0.592±0.023**	**0.591±0.030**	**0.562±0.029**	**0.628±0.018**	**0.599±0.017**	**0.609±0.022**	**0.567±0.024**
				AUPR↑			
CREM-C	0.824±0.111	0.849±0.033	0.801±0.066	0.853±0.039 •	0.846±0.063 •	0.750±0.017 •	0.572±0.032 •
CREM-K	0.839±0.078	0.875±0.027	0.702±0.009 •	0.894±0.029 •	0.876±0.050 •	0.789±0.010	0.639±0.023
CREM	**0.840±0.077**	**0.877±0.028**	**0.806±0.046**	**0.898±0.028**	**0.879±0.048**	**0.793±0.013**	**0.648±0.023**

Table 5: Experimental results of CREM and variants (mean±std) in terms of *AUROC* and *AUPR*. ↑ (↓) indicates the larger (smaller) the value, the better the performance. The best and second best results are highlighted in **boldface** and underline, respectively. In addition, •/◦ indicates whether CREM is statistically superior/inferior to the comparing approaches on each data set with pairwise t-test (at 0.05 significance level).

- CREM-K is implemented by learning reciprocal points directly in the kernel space, which is the degenerate case where the classifier-induced mechanism is excluded.

Table 5 shows the detailed experimental results for CREM against its variants on each evaluation metric and Table 6 summarizes the win/tie/loss counts (pairwise t-test at 0.05 significant level). Compared with the two variants, we can observe CREM achieves statistically superior performance against them in terms of each evaluation metric, demonstrating the effectiveness of the two essential components in CREM for reciprocal points induction.

Metrics	CREM **against**	
	CREM-C	CREM-K
AUROC	6/1/0	3/4/0
AUPR	4/3/0	3/4/0
In Total	**10/4/0**	**6/8/0**

Table 6: Win/tie/loss counts (pairwise t-test at 0.05 significant level) for CREM against its variants.

Complexity Analysis. In CREM, the main computational cost lies in two critical procedures: (1) the construction of kernel matrix $\mathbf{K}$ and the similarity matrix $\mathbf{C}$; (2) the alternating optimization process. The training complexity of the first procedure is $\mathcal{O}(m^2d + q^2)$. For the latter, the training complexity of one iteration is $\mathcal{O}(qm^2 + m^3)$. Due to the fact that the former procedure is executed only once, the overall training complexity of CREM is approximately equivalent to the complexity of the latter procedure. The detailed empirical training and test time of each comparing approach are reported in Appendix C, which shows that the time overhead of CREM is comparable to those of existing approaches.

Sensitivity Analysis. Figure 1 illustrates how the performance of CREM varies with different values of λ_1 and λ_2 mentioned in Eq.(11). The other parameters are fixed according to the settings in the Configuration paragraph. It is shown that the performance of CREM tends to underperform when λ_1 and λ_2 increase. Since λ_1 and λ_2 control the model complexity and label correlations respectively, CREM might suffer from underfitting issue if λ_1 and λ_2 are too large. Similar results can be found in other cases. Therefore, we take $\lambda_1 = 1$ and $\lambda_2 = 0.1$ as the fixed parameters in this paper. We also perform sensitivity analyses for λ_3, α and K, which can be found in Appendix B.

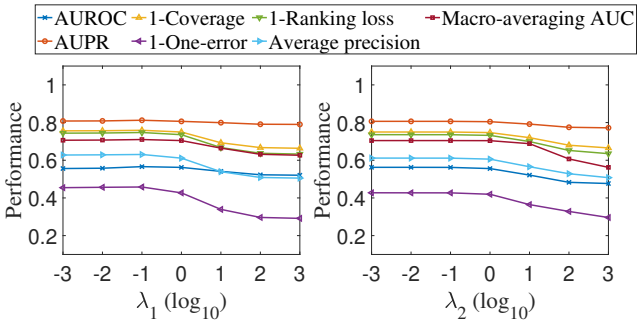

Figure 1: Performance of CREM with varying value of trade-off parameters λ_1 and λ_2 on slashdot.

Conclusion

In this paper, we present the first attempt towards leveraging reciprocal points to address the MLOSR problem. In contrast to the existing method that primarily tackles this challenging problem through sub-labeling approximations and prototype-based comparisons, we propose a novel approach CREM, which explicitly incorporates unknown label information into the learning process. Specifically, the implicit unknown label information is formulated via the distance relations with reciprocal points, allowing the open space risk to be indirectly managed by instances with known labels in a learnable margin. After that, we further utilize an RBF kernel function to eliminate the instance-dependent bias, and introduce a unified optimization framework to jointly facilitate the classifier and the induction of reciprocal points. Comprehensive experiments demonstrate the effectiveness and superiority of our proposed CREM approach. In the future, it will be interesting to incorporate multi-label metric learning into the induction procedure of reciprocal points, and extract discriminative feature representations from large-scale image datasets to enhance the robustness of the proposed paradigm.

Acknowledgments

The authors wish to thank the anonymous reviewers for their helpful comments and suggestions. This work was supported by the National Science Foundation of China (62225602), the SEU Innovation Capability Enhancement Plan for Doctoral Students (CXJH_SEU 25130), and the Big Data Computing Center of Southeast University.

References

Boutell, M. R.; bo Luo, J.; Shen, X.-P.; and Brown, C. M. 2004. Learning multi-label scene classification. *Pattern Recognition*, 37(9): 1757–1771.

Brinker, C.; Mencía, E. L.; and Fürnkranz, J. 2014. Graded multilabel classification by pairwise comparisons. In *Proceedings of the 14th IEEE International Conference on Data Mining*, 731–736. Shenzhen, China.

Chen, G.; Peng, P.; Wang, X.; and Tian, Y. 2021. Adversarial reciprocal points learning for open set recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(11): 8065–8081.

Chen, G.; Qiao, L.; Shi, Y.; Peng, P.; Li, J.; Huang, T.; Pu, S.; and Tian, Y. 2020. Learning open set network with discriminative reciprocal points. In *Proceedings of the 16th European conference on computer vision*, 507–522. Glasgow, UK.

Chen, W.; Mao, J.-X.; and Zhang, M.-L. 2025. Learnware Specification via Label-Aware Neural Embedding. In *Proceedings of the 39th AAAI Conference on Artificial Intelligence*, volume 39, 15857–15865. Philadelphia, PA.

Da, Q.; Yu, Y.; and Zhou, Z.-H. 2014. Learning with augmented class by exploiting unlabeled data. In *Proceedings of the 28th AAAI conference on artificial intelligence*, volume 28, 1760–1766. Québec, Canada.

Demsar, J. 2006. Statistical Comparisons of Classifiers over Multiple Data Sets. *J. Mach. Learn. Res.*, 7: 1–30.

Fürnkranz, J.; Hüllermeier, E.; Mencía, E. L.; and Brinker, K. 2008. Multilabel classification via calibrated label ranking. *Machine Learning*, 73(2): 133–153.

Gao, Y.; Xu, M.; and Zhang, M. 2023. Unbiased Risk Estimator to Multi-Labeled Complementary Label Learning. In *Proceedings of the 32th International Joint Conference on Artificial Intelligence*, 3732–3740. Macao, SAR, China.

Goodfellow, I. J.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; and Bengio, Y. 2014. Generative adversarial nets. In *Advances in neural information processing systems 27*, volume 27, 2672–2680. Quebec, Canada.

Gu, Z.-Z.; Jia, B.-B.; and Zhang, M.-L. 2026. Similarity-based multi-dimensional multi-label classification. *Frontiers of Computer Science*, 20(2): Article 2002322.

Guo, Y.; Camporese, G.; Yang, W.; Sperduti, A.; and Ballan, L. 2021. Conditional variational capsule network for open set recognition. In *Proceedings of the 18th IEEE/CVF International Conference on Computer Vision*, 103–111. QC, Canada.

Hang, J.-Y.; and Zhang, M.-L. 2024a. Binary decomposition: A problem transformation perspective for open-set semi-supervised learning. In *Proceedings of the 41st International Conference on Machine Learning*. Vienna, Austria.

Hang, J.-Y.; and Zhang, M.-L. 2024b. Dual perspective of label-specific feature learning for multi-label classification. *ACM Transactions on Knowledge Discovery from Data*, 19(1): 1–30.

Huang, M.; Zhuang, F.-Z.; Zhang, X.; Ao, X.; Niu, Z.-Y.; Zhang, M.; and He, Q. 2019. Supervised representation learning for multi-label classification. *Machine Learning*, 108(5): 747–763.

Jia, B.-B.; Liu, J.-Y.; Hang, J.-Y.; and Zhang, M.-L. 2023. Learning label-specific features for decomposition-based multi-class classification. *Frontiers of Computer Science*, 17(6): 176348.

Kou, Z.; Wang, J.; Tang, J.; Jia, Y.; Shi, B.; and Geng, X. 2024. Exploiting Multi-Label Correlation in Label Distribution Learning. In *Proceedings of the Thirty-Third International Joint Conference on Artificial Intelligence*, 4326–4334.

Liu, C.; Yang, C.; Qin, H.-B.; Zhu, X.; Liu, C.-L.; and Yin, X.-C. 2023. Towards open-set text recognition via label-to-prototype learning. *Pattern Recognition*, 134: 109109.

Liu, F. T.; Ting, K. M.; and Zhou, Z. 2008. Isolation Forest. In *Proceedings of the 8th IEEE International Conference on Data Mining*, 413–422. Pisa, Italy.

Liu, W.-W.; Wang, H.-B.; Shen, X.-B.; and Tsang, I. W. 2021. The emerging trends of multi-label learning. *IEEE transactions on pattern analysis and machine intelligence*, 44(11): 7955–7974.

Ma, J.-S.; and Perkins, S. 2003. Time-series novelty detection using one-class support vector machines. In *Proceedings of the 2003 International Joint Conference on Neural Networks*, volume 3, 1741–1745. Portland, OR.

Mao, J.; Wang, W.; and Zhang, M. 2023. Label Specific Multi-Semantics Metric Learning for Multi-Label Classification: Global Consideration Helps. In *Proceedings of the 32th International Joint Conference on Artificial Intelligence*, 4055–4063. Macao, SAR, China.

Mao, J.-X.; Hang, J.-Y.; and Zhang, M.-L. 2024. Learning label-specific multiple local metrics for multi-label classification. In *Proceedings of the 33rd International Joint Conference on Artificial Intelligence*, 4742–4750. Jeju, South Korea.

Mao, J.-X.; Rui, Y.; and Zhang, M.-L. 2025. Implicit Relative Labeling-Importance Aware Multi-Label Metric Learning. In *Proceedings of the 39th AAAI Conference on Artificial Intelligence*, volume 39, 19414–19422. Philadelphia, PA.

McAuley, J.; Pandey, R.; and Leskovec, J. 2015. Inferring networks of substitutable and complementary products. In *Proceedings of the 21th ACM SIGKDD international conference on knowledge discovery and data mining*, 785–794. Sydney, Australia.

Neal, L.; Olson, M.; Fern, X.; Wong, W.-K.; and Li, F. 2018. Open set learning with counterfactual images. In *Proceedings of the 15th European conference on computer vision*, 613–628. Munich, Germany.

Oza, P.; and Patel, V. M. 2019. C2ae: Class conditioned auto-encoder for open-set recognition. In *Proceedings of the 30th IEEE/CVF conference on computer vision and pattern recognition*, 2307–2316. Long Beach, CA.

Scheirer, W. J.; de Rezende Rocha, A.; Sapkota, A.; and Boult, T. E. 2013. Toward Open Set Recognition. *IEEE Trans. Pattern Anal. Mach. Intell.*, 35(7): 1757–1772.

Schölkopf, B.; and Smola, A. J. 2002. *Learning with kernels: support vector machines, regularization, optimization, and beyond*. MIT press.

Shi, J.; Wei, T.; and Li, Y. 2024. Residual diverse ensemble for long-tailed multi-label text classification. *Science China Information Sciences*, 67(11): Article 212102.

Sun, X.; Yang, Z.; Zhang, C.; Ling, K.-V.; and Peng, G. 2020. Conditional gaussian distribution learning for open set recognition. In *Proceedings of the 31th IEEE/CVF conference on computer vision and pattern recognition*, 13480–13489. Seattle, WA.

Tang, P.; Jiang, M.; Xia, B. N.; Pitera, J. W.; Welser, J.; and Chawla, N. V. 2020. Multi-Label Patent Categorization with Non-Local Attention-Based Graph Convolutional Network. In *Proceedings of the 34th AAAI conference on artificial intelligence*, 9024–9031. New York, NY.

Wang, Y.; Hang, J.; and Zhang, M. 2022. Stable Label-Specific Features Generation for Multi-Label Learning via Mixture-Based Clustering Ensemble. *IEEE CAA J. Autom. Sinica*, 9(7): 1248–1261.

Wang, Y.; Mu, J.; Zhu, P.; and Hu, Q. 2024. Exploring diverse representations for open set recognition. In *Proceedings of the 36th AAAI Conference on Artificial Intelligence*, volume 38, 5731–5739. Vancouver, Canada.

Wang, Y.-B.; Hang, J.-Y.; and Zhang, M.-L. 2024. Multi-label open set recognition. In *Advances in Neural Information Processing Systems 37*, volume 37, 5739–5756. Vancouver, Canada.

Wang, Z.; Xu, Q.; Yang, Z.; He, Y.; Cao, X.; and Huang, Q. 2022. Openauc: Towards auc-oriented open-set recognition. In *Advances in Neural Information Processing Systems 35*, volume 35, 25033–25045. New Orleans, LA.

Wei, Q.; Dobigeon, N.; and Tourneret, J.-Y. 2015. Fast fusion of multi-band images based on solving a Sylvester equation. *IEEE Transactions on Image Processing*, 24(11): 4109–4121.

Wei, T.; Shi, J.; and Li, Y. 2021. Probabilistic Label Tree for Streaming Multi-Label Learning. In *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, 1801–1811. Virtual Event, Singapore.

Xu, B.; Shen, F.; and Zhao, J. 2023. Contrastive open set recognition. In *Proceedings of the 35th AAAI conference on artificial intelligence*, volume 37, 10546–10556. Washington, DC.

Yang, H.-M.; Zhang, X.-Y.; Yin, F.; Yang, Q.; and Liu, C.-L. 2020. Convolutional prototype network for open set recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(5): 2358–2370.

You, R.-C.; Guo, Z.-Y.; Cui, L.; Long, X.; Bao, Y.-Z.; and Wen, S.-L. 2020. Cross-modality attention with semantic graph embedding for multi-label classification. In *The 34th AAAI Conference on Artificial Intelligence*, 12709–12716. New York, NY.

Yuan, M.; and Wegkamp, M. 2010. Classification Methods with Reject Option Based on Convex Risk Minimization. *Journal of Machine Learning Research*, 11(1).

Zhang, M.; Li, Y.; Liu, X.; and Geng, X. 2018. Binary relevance for multi-label learning: An overview. *Frontiers of Computer Science*, 12(2): 191–202.

Zhang, M.; and Wu, L. 2015. Lift: Multi-Label Learning with Label-Specific Features. *IEEE Trans. Pattern Anal. Mach. Intell.*, 37(1): 107–120.

Zhang, M.-L.; and Zhou, Z.-H. 2013. A review on multi-label learning algorithms. *IEEE transactions on knowledge and data engineering*, 26(8): 1819–1837.

Zhou, D.-W.; Ye, H.-J.; and Zhan, D.-C. 2021. Learning placeholders for open-set recognition. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 4401–4410. Virtual Conference.

Zhu, Y.; Ting, K. M.; and Zhou, Z. 2018. Multi-Label Learning with Emerging New Labels. *IEEE Trans. Knowl. Data Eng.*, 30(10): 1901–1914.