Pattern Recognition 150 (2024) 110358

Contents lists available at ScienceDirect

Pattern Recognition

journal homepage: www.elsevier.com/locate/pr

Semi-supervised imbalanced multi-label classification with label propagation

Guodong Du [a,b], Jia Zhang [c,*], Ning Zhang [a], Hanrui Wu [c], Peiliang Wu [a,b], Shaozi Li [d,e]

[a] School of Information Science and Engineering, Yanshan University, Qinhuangdao, 066004, PR China
[b] The Key Laboratory for Computer Virtual Technology and System Integration of Hebei Province, Yanshan University, Qinhuangdao, 066004, PR China
[c] College of Information Science and Technology, Jinan University, Guangzhou 510632, PR China
[d] Department of Artificial Intelligence, Xiamen University, Xiamen, 361005, PR China
[e] Fujian Key Laboratory of Big Data Application and Intellectualization for Tea Industry, Wuyi University, Wuyishan, 354300, PR China

ARTICLE INFO

Keywords:
Label propagation
Class-imbalance learning
Semi-supervised multi-label learning
Collaboration technique

ABSTRACT

Multi-label learning tasks usually encounter 遇到 the problem of the class-imbalance, where samples and their corresponding labels are non-uniformly distributed over multi-label data space. It has attracted increasing attention during the past decade, however, there is a lack of methods capable of handling the imbalanced problem in a semi-supervised setting. This study proposes a label propagation 标签传播 technique to settle the semi-supervised imbalanced multi-label issue. Specially, we first utilize a collaborative manner to exploit the 利用标签和实例的关联性 correlations from labels and instances, and learn a label regularization matrix to overcome the imbalanced problem in the labeled instance. After that, we extend to semi-supervised learning and explore to represent the similarity of instances with weighted graphs on labeled and unlabeled data. Then, the data distribution information and label correlations are fully utilized to design the loss function under the consistency assumption manner. At last, we present an iterative scheme to settle the optimization issue, thereby achieving label propagation to address the imbalanced challenge. Experiments on a variety of multi-label data sets show the favorable performance of the proposed method against related comparing approaches. Notably, the proposed method is also validated to be robust with a limited number of training instances.

1. Introduction

Multi-label learning handles the example which may be associated with multiple labels simultaneously [1]. Over the last decade, multi-label learning has various applications in the real world, such as automatic image annotation [2], disease diagnosis [3], and gene expression prediction [4]. For instance, an image can be annotated with multiple scenes, such as clouds, city, mountain, sky, sun, and tree [5]; a coronavirus disease 2019 (COVID-19) patient may be tagged with multiple patterns of syndromes, such as fever, fatigue, cough, nausea and vomiting, and dyspnea [6]; a gene could be associated with several functional classes, such as ionic homeostasis, cellular biogenesis, protein synthesis, transcription, and metabolism energy [7]. As an instance has multiple labels, the class distribution of each label can be highly imbalanced [8]. Fig. 1 provides an example of imbalanced multi-label issue on the Enron data set (in Table 2). Intuitively, it is easy to see that the imbalanced multi-label issue comes from two aspects, the one is that the amount of negative labels of each instance is far more than its positive labels, and another is that the amount of positive instances of each label is significantly different. Hence, dealing with large and imbalanced datasets where finding a relevant subset of labels for an object can be challenging and computationally expensive. Thus, the class-imbalance is a vital problem in multi-label learning [9]. Recently imbalanced multi-label learning has gained increasing interest, which tries to exploit the inherent 固有的 property of labels, e.g., label correlations, to decrease the difficulty of classification [10]. Nevertheless, the performance degradation 降解 caused by the imbalanced issue remains to be a huge challenge that is not solved well.

In addition, existing methods for multi-label learning mainly focus on a full supervised manner. However, it is expensive and difficult to acquire labeled training instances in real-world applications [11], 特别 particularly training data associated with a large number of labels, thus resulting in only a minor quantity of labeled data that can be utilized for the modeling. For example, in gene expression prediction task, labeled genomics data are extremely scarce [4]. Similarly, in traditional Chinese medicine (TCM) state identification task, the labeled syndromes of patients are also extremely limited [12]. As a result, the labeled data probably are insufficient for multi-label learning. Lacking supervised information of unlabeled data makes it hard to predict their instances correctly [13]. Considering the decrease of learning

* Corresponding author.
E-mail addresses: gddu@ysu.edu.cn (G. Du), jiazhang@jnu.edu.cn (J. Zhang), znrx@stumail.ysu.edu.cn (N. Zhang), hrwu@outlook.com (H. Wu), peiliangwu@ysu.edu.cn (P. Wu), szlig@xmu.edu.cn (S. Li).

https://doi.org/10.1016/j.patcog.2024.110358
Received 9 July 2023; Received in revised form 21 January 2024; Accepted 19 February 2024
Available online 21 February 2024

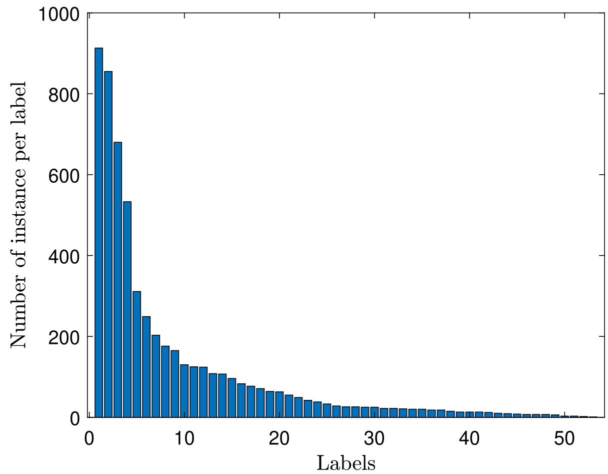


Fig. 1. Example for imbalanced multi-label problem on Enron data set (in this illustration, the label rank is set as the number of instances per label).

performance caused by limited annotated data, it should be noted that a promising direction to improve the learning performance requires that we exploit the unlabeled data [14]. At present, there are plenty of studies on semi-supervised multi-label learning. Semi-supervised learning is suitable for learning with both labeled and unlabeled data simultaneously. For instance, many researchers devote to utilizing 标签传播 label propagation techniques [15], 共同训练 co-training methods [16] and manifold regularization based approaches [17] to achieve semi-supervised multi-label learning. Among these approaches, a widely used way to solve this issue is with label propagation. Label propagation can increase classification performance by exploiting the structure of unlabeled data and incorporating them into the classification process [18]. However, the existing label propagation based methods are usually incapable of exploiting label dependencies [19], more importantly, 没有考虑类别不平衡 do not consider the class-imbalance issue for multi-label learning. Thus, they are prone to propagate class imbalanced outcomes, and may deteriorate overall performance.

To settle the above problems, this study presents a semi-supervised imbalanced multi-label learning approach with label propagation, called SMCLP. 1.同时在标签空间和特征空间研究标签之间的关联性和实例之间的关联性 SMCLP first explores both the label correlations in label space and the instance similarity in feature space simultaneously in a collaborative manner. 2. 计算标签关联矩阵，并将其融入到loss中，用于克服不平衡问题 Most importantly, SMCLP calculates a label regularization matrix from the labeled instance and incorporates it into the loss function to overcome the imbalanced issue. In order to obtain better semi-supervised learning performance, SMCLP also deploys a 在有标签数据和无标签数据计算加权无向图用于编码不同实例之间的相似性 weighted undirected graph on labeled and unlabeled data to encode the similarity between all instances. Finally, SMCLP provides a loss function under the consistency assumption to enhance the training performance, and can be solved efficiently to implement the imbalanced problem. Extensive experiments reveal that this study can achieve a promising performance against related well-established approaches. In summary, the major contributions of this study are as follows:

- We propose a new semi-supervised imbalanced multi-label classification method, which pays special attention to the inherent property of class-imbalance in multi-label data, and utilizes label propagation to assist in the implementation.
- For class-imbalance learning, we utilize a label regularization matrix to handle the imbalanced multi-label problem and leverage a collaborative manner to ensure the balanced outcomes; For semi-supervised learning, we exploit the representation of labeled and unlabeled data through encoding the similarity between all instances with weighted graph.
- An alternating minimization strategy is used to search for the optimal solution of the proposed method.
- Extensive studies are carried out on ten multi-label data sets to validate the effectiveness of the proposed method.

The remainder of this study is structured as follows. In Section 2, we provide a detailed review of multi-label learning, imbalanced multi-label learning, and semi-supervised multi-label learning methods. The proposed method is detailed in Section 3. The experimental results are presented in Section 4. At last, we summarize this study and point out future research directions in Section 5.

2. Related work

Over the last decade, numerous approaches have been presented to improve the multi-label classification performance [20]. The relatively straightforward multi-label learning approach splits the issue into several individual binary classification tasks, one for each label [21]. However, this method is easy to accomplish, but it probably causes a decline in performance due to the neglect of label correlations. To solve the above issue, exploring label correlations has been commonly regarded as a crucial part for effective multi-label learning [1]. In terms of the order of label correlations being considered, most existing multi-label learning methods can be decomposed into three groups: first-order methods, second-order methods, and high-order methods [22]. For the first-order methods, which ignore label correlation, the multi-label issue is transformed into various single-label classification tasks. Binary relevance (BR) [23] and multi-label k-nearest neighbor method (ML-KNN) [24] are representative ones of this class. For the second-order methods, the multi-label issue can be solved by exploring the correlations between label pairs. Some representative methods contain backpropagation for multi-label learning (BP-MLL) [25], calibrated label ranking (CLR) [26], and ranking support vector machine (Rank-SVM) [7]. For the third group, high-order methods mine the correlations from all the labels or a subset of labels. Following this principle, many high-order methods have been put forward, including classifier chains (CC), ensembles of CC (ECC) [27], random k-labelsets (RAKEL) [28], and more.

A commonness of the aforementioned methods is that they deal with multi-label learning issue by manipulating label correlations [29]. Exploiting the label correlation is a key technology for multi-label learning [30]. The performance decline caused by another intrinsic property, e.g., class-imbalance, is a major problem that cannot be ignored. Over the last decade, plenty of class-imbalance learning approaches have been presented. According to the categorized criterion of imbalanced learning approaches on single-label data sets [31], these imbalanced multi-label learning approaches can be categorized into three groups, including data sampling methods, cost-sensitive methods, and ensemble learning methods [9]. The data sampling methods exhibit comprehensive attention towards equitably distributing the number of samples across different classes within the dataset before an effective classification model. For example, the multi-label synthetic minority over-sampling technique (MLSMOTE) [32] exploited ranking as the label generation method to decrease the level of imbalance in multi-label data sets. In contrast, the cost-sensitive approaches rely on the cost matrix values for the effective implementation of class-imbalance learning. For instance, CPNL [33] attempted to address the class-imbalance problem by leveraging the correlations between positive and negative labels in a pairwise manner. The ensemble learning approaches are also employed to tackle the challenge of imbalanced multi-label. One well-known approach is COCOA [10], which has taken into account the randomness in pairwise coupling and made use of ensemble learning to aggregate the predictions of randomly generated imbalanced learners. Note that there are also some approaches studied on new strategies for imbalanced multi-label learning. For instance, MMIB [34] tackled the class-imbalance problem by imposing label consistency via submodular minimization. GCLE [35] aimed to solve the class-imbalance issue in both single-label and multi-label data sets by introducing graph constrains. Bal-2D-MALIU [36] extended the version space theory from single-label data to multi-label data and considered class-imbalance issue to improve the multi-label classification performance. MKML [37] aimed to tackle two challenges of label relation and class imbalance in multi-label learning by involving multiple kernel learning strategy. In addition, ML-CIB [38] utilized label regularization via accelerated proximal gradient to achieve imbalanced multi-label learning with incomplete label space.

All of the above mentioned methods have one common hypothesis, that is, there have a large quantity of labeled data. Nevertheless, considering the high cost of data annotation, practical applications often require the processing of a large amount of unlabeled data, resulting in a relatively small proportion of labeled data [39]. This poses a significant challenge for machine learning algorithms, since labeled data are key requirement for effective training and optimal performance. Semi-supervised learning is a common framework in machine learning that aims at learning from both labeled and unlabeled data. Taking advantage of unlabeled data along with labeled data frequently performs better results than using the labeled data alone. In a general, semi-supervised multi-label learning is decomposed into two groups, namely inductive multi-label learning and transductive multi-label learning [40]. Inductive multi-label learning is pure semi-supervised multi-label learning that predicts test data through a trained model [41]. iMLCU [42] is the first attempt towards an inductive manner to perform semi-supervised multi-label learning. Moreover, COINS [16] used the famous co-training method to construct two classification models by combining the outputs to yield the final predictions. Transductive multi-label learning assumes that test data are derived from unlabeled data, and the aim is to acquire the optimal solution on the test data [43]. One representative method is the TRAM [44], which constructed a similarity graph among labeled and unlabeled data, and then predicted the labels of unlabeled data via label propagation. ML-RMG [45] developed an ensemble learning method, which integrates multi-label learning and graph-based semi-supervised learning into one framework. MGLP [46] utilized multi-level neighborhood information granularity and a three-way decision method, where the three-way decision method can be used to select unlabeled data for further annotation. Besides, CMLP [47] tackled the semi-supervised multi-label learning issue by propagating the independent label information to recover the original labels in a collaborative manner. Moreover, there still exist some semi-supervised approaches to resolve multi-label learning with incomplete labels. For instance, SSWL [48] applied the label similarity and the instance similarity to the complement of missing labels, and then established an optimization framework to accomplish weak-label classification through an ensemble of multiple models. MSWL [41] utilized the manifold regularized sparse model for learning from weakly labeled data.

It is worth mentioning that the proposed method is most relevant to collaboration based multi-label learning (CAMEL) [49]. To be specific, CAMEL is guided by collaborative assumption, that is, the final prediction of a label composes the collaboration between its prediction and the prediction of others, thus achieving multi-label learning in a supervised learning scenario. Distinguished from the assumptions of CAMEL, we exploit a label regularization matrix to handle the imbalanced multi-label problem. Most importantly, we consider to explicitly explore the label correlation and instance correlation in a collaborative manner and achieve label propagation in a semi-supervised setting. For that, we encode the correlations from labels and instances, and learn the representation of labeled and unlabeled data utilizing a weighted graph to achieve excellent results on a variety of data sets.

3. The SMCLP algorithm

Formally, let $\boldsymbol{X} \in \mathbb{R}^{n\times m}$ be a data set with labeled and unlabeled data, where $\boldsymbol{x}_i \in \mathcal{X}$ stands for the ith instance comprising m features, and n denotes the total number of instances. The target of labeled data is represented by $\boldsymbol{Y} \in \mathbb{R}^{l\times q}$, where l denotes the number of labeled data, typically $l \ll n$, and $\boldsymbol{y}_i \in \{y_{i1}, y_{i2}, \ldots, y_{iq}\}$ is accompanying with a finite set of q possible labels. For $y_{ij} \in \boldsymbol{y}_i$, $y_{ij} = 1$ in case that jth label is related to $\boldsymbol{x}_i$, otherwise $y_{ij} = -1$. Thus, the goal of this work is to study a predictive label matrix $\widehat{\boldsymbol{Y}} \in \{1, -1\}^{n\times q}$ from $\{\boldsymbol{X}, \boldsymbol{Y}\}$ with a well-designed label propagation method. In order to make the model description easy to understand, the frequently used notations are illustrated in Table 1.

Based on the above mentioned setting, we describe the proposed method for handling the imbalanced multi-label issue under semi-supervised setting, as illustrated in Fig. 2. More specifically, we first utilize the information of label correlations and data distribution in a collaborative manner. Meanwhile, we also estimate a label regularization matrix, and incorporate it into the loss function to handle the imbalanced multi-label issue. After that, with the help of instance similarity, the weighted graph is constructed. Then, we develop the loss function under the consistency assumption to enhance the training performance. Finally, a classifier is trained and combined with the proposed optimized solution to predict unlabeled data, thus achieving the purpose. We will illustrate key modules of the proposed method in detail.

3.1. Collaboration based imbalanced multi-label learning

3.1.1. Label collaboration

In general, vanilla label propagation methods fail to make use of the label correlations. To explore the label correlations in the label propagation process, we can achieve the final label from the perspective of collaborative assumption. Most importantly, we utilize a collaborative manner to exploit the correlations from labels, and learn an 中间的 intermediate variable $\boldsymbol{Z} \in \mathbb{R}^{l\times q}$ for label propagation. Suppose $\boldsymbol{Z}$ is a fitting label space induced from the predictive model. Therefore, the final prediction of a label is comprised of two parts: its own prediction part and the prediction of other labels (collaborative part). Hence, we can formulate the prediction results of label prediction as

$$\phi(\boldsymbol{Z}) = (1-\alpha)\boldsymbol{Z} + \alpha \boldsymbol{Z}\boldsymbol{S}, \tag{1}$$

where α is a trade-off parameter to reflect the collaboration degree, $\boldsymbol{S} \in \mathbb{R}^{q\times q}$ denotes correlation matrix, $s_{ij} \in \boldsymbol{s}_i$ is reflect the influence of the i-label over the j-label. Moreover, a label is relevant to only a few labels. Thus, the relationships between a label with others should be sparse. Therefore, we can obtain the following optimization issue:

$$\begin{aligned} &\min_{\boldsymbol{s}_i} \left\| \left((1-\alpha)\boldsymbol{y}_i + \alpha \boldsymbol{Y}\boldsymbol{s}_i\right) - \boldsymbol{y}_i \right\|^2 + \gamma \left\|\boldsymbol{s}_i\right\|^2 \\ &\text{s.t. } s_{ii} = 0, \end{aligned} \tag{2}$$

where γ is a regularization parameter. The optimization issue can be converted into an ordinary ridge regression problem, and can be efficiently solved.

In addition, most of multi-label learning methods do not consider the class-imbalance problem, and assume balanced relevant and irrelevant samples in each label. However, there are two aspects that need to be considered: instance class imbalance and label class imbalance. For instance class imbalance, we need to take full account of the most of the labeled instances that have much fewer relevant samples than the irrelevant ones. For label class imbalance, we need to consider that the amount of relevant samples varies significantly across labels. Moreover, we also need to consider the balanced outcomes when performing the label propagation process. Therefore, we propose to utilize the label correlation to learn a label regularization matrix $\boldsymbol{K}$, which controls the influence of labels on the labeled instance. The matrix $\boldsymbol{K}$ takes advantage of label correlation to modify the predictive confidence values. By considering the label correlations, the matrix $\boldsymbol{K} = \left[k_{ij}\right]_{q\times q}$ is given by the target of labeled instance $\boldsymbol{Y}$ as

$$k_{ij} = \frac{\sum_{i,j,u=1}^{l} \Omega_{Y_{iu}}\left(Y_{ju}\right)}{l} \quad \text{for } i = 1\cdots q, j = i+1, \tag{3}$$

Table 1
List of the frequently used notations in this study.

Notations	Explanation	Notations	Explanation
$\boldsymbol{X}$	Feature matrix	$\boldsymbol{Y}$	Label matrix
n	Number of instances	m	Number of features
l	Number of labeled instances	q	Number of labels
$\hat{\boldsymbol{Y}}$	Predictive label matrix	$\boldsymbol{G}$	Undirected graph
$\boldsymbol{E}$	Edges set	$\boldsymbol{V}$	Vertex set
$\boldsymbol{W}$	Non-negative weight matrix	σ	Parameter of similarity calculation
θ	Smoothing factor of sample's weight	$\boldsymbol{P}$	Propagation matrix
$\boldsymbol{D}$	Diagonal matrix of $\boldsymbol{W}$	$\boldsymbol{F}$	Optimal prediction label matrix
$\boldsymbol{Z}$	Intermediate variable to the labeled data	$\boldsymbol{S}$	Correlation matrix of label matrix
$\boldsymbol{C}$	Correlation matrix of instance matrix	α and δ	Reflect the collaboration degree of labels and instances
$\boldsymbol{Q}$ and $\boldsymbol{B}$	Intermediate variable of correlation matrix	γ and τ	Regularization parameters
μ, λ and ζ	Trade-off parameter for objective function		

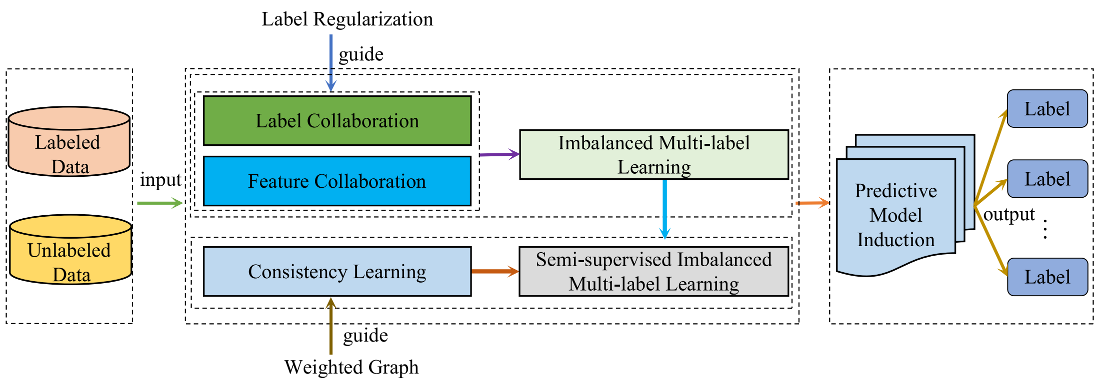


Fig. 2. Flowchart of the proposed method.

where $\Omega_{Y_{iu}}\left(Y_{ju}\right) = \begin{cases} 1 & \text{if } Y_{iu} = Y_{ju} \\ 0 & \text{if } Y_{iu} \neq Y_{ju} \end{cases}$. Thus, we consider the class-imbalance problem, and use the label regularization matrix $\boldsymbol{K}$ to control the influence of labels. The corresponding optimization problem is formulated as

$$\mathcal{L}_{lab}(\boldsymbol{Z}) = \frac{1}{2}\|\phi(\boldsymbol{Z}) - \boldsymbol{Y}\boldsymbol{K}\|_{\mathcal{F}}^2, \tag{4}$$

where $\|\cdot\|_{\mathcal{F}}$ stands for the Frobenius norm.

3.1.2. Feature collaboration

Guided by the collaborative assumption, we can extend the collaborative relationship between labels to feature space. Therefore, the label of each unlabeled data is also influenced by the feature of itself and its nearby samples. Hence, we can formulate the prediction result as

$$\varrho(\boldsymbol{Z}) = (1-\delta)\boldsymbol{Z} + \delta\boldsymbol{C}\boldsymbol{Z}, \tag{5}$$

where δ is a trade-off parameter to reflect the collaboration degree, $\boldsymbol{C} \in \mathbb{R}^{l\times l}$ is the correlation matrix, $c_{ij} \in \mathbf{c}_i$ is reflect the influence of the ith instance over the jth instance in feature space. In a similar way, the correlation matrix $\boldsymbol{C}$ can be easily estimated by the following formula

$$\begin{aligned} &\min_{\mathbf{c}_i} \left\|\left((1-\delta)\mathbf{x}_i^l + \delta\mathbf{c}_i\mathbf{X}^l\right) - \mathbf{x}_i^l\right\|^2 + \tau\left\|\mathbf{c}_i\right\|^2 \\ &\text{s.t. } c_{ii} = 0, \end{aligned} \tag{6}$$

where $\mathbf{X}^l = \left[\mathbf{x}_1^l, \mathbf{x}_2^l, \ldots, \mathbf{x}_l^l\right]^\top \in \mathbb{R}^{l\times m}$ is the labeled instance, and τ is a regularization parameter. Similarly, this optimization problem also can transform into a ridge regression problem and achieve the solution immediately. Therefore, we can obtain a new regularization term:

$$\mathcal{L}_{fea}(\boldsymbol{Z}) = \frac{1}{2}\|\varrho(\boldsymbol{Z}) - \boldsymbol{Y}\boldsymbol{K}\|_{\mathcal{F}}^2, \tag{7}$$

As stated above, we consider the class-imbalance problem in multi-label learning, and leverage the collaborative information from label correlations and data distribution simultaneously. Therefore, we can obtain the following optimization framework:

$$\mathcal{L}(\boldsymbol{Z}) = \frac{\zeta}{2}\|\phi(\boldsymbol{Z}) - \boldsymbol{Y}\boldsymbol{K}\|_{\mathcal{F}}^2 + \frac{\lambda}{2}\|\varrho(\boldsymbol{Z}) - \boldsymbol{Y}\boldsymbol{K}\|_{\mathcal{F}}^2 \tag{8}$$

where ζ and λ are trade-off parameters, the fitting label space $\boldsymbol{Z}$ works as a bridge between imbalanced multi-label learning and label propagation.

3.2. Extension to semi-supervised learning

As the fitting label space $\boldsymbol{Z}$ is ready, we can utilize it to achieve label propagation. Thus, we leverage the labeled and unlabeled instances to construct the weighted graph and develop the loss function under the consistency assumption to enhance the semi-supervised imbalanced multi-label learning performance.

3.2.1. Weighted graph construction

According to the smoothness assumption, the same class label should be shared by similar samples on the graph. Thus, we can predict the unlabeled samples through the graph. However, it is exceptionally difficult to construct high quality graphs. The essential task of graph construction is to measure the similarities between samples. Motivated by the above perspective, we present a weighted graph construction strategy to improve the quality and robustness of the constructed graph. Therefore, we define a undirected graph $\boldsymbol{G} = \langle \boldsymbol{E}, \boldsymbol{V}, \boldsymbol{W}\rangle$ on $\mathcal{X}$, in which $\boldsymbol{E}$ denotes the edges set, $\boldsymbol{V}$ stands for the vertex set, and $\boldsymbol{W}$ is the non-negative weight matrix. Unlike vanilla label propagation approaches, the proposed method considers the weights of each edge and adjusts the weights to capture the most discriminative data characteristics. Thus, we define $\boldsymbol{W} = \left[w_{ij}\right]_{n\times n}$ as an affinity matrix, and

$$w_{ij} = \begin{cases} \left(sim(\mathbf{x}_i, \mathbf{x}_j)\right)^\theta & \mathbf{x}_j \in \mathcal{O}\left(\mathbf{x}_i\right) \text{ or } \mathbf{x}_i \in \mathcal{O}\left(\mathbf{x}_j\right) \\ 0 & \text{otherwise} \end{cases}, \tag{9}$$

in which $sim(\mathbf{x}_i, \mathbf{x}_j) = \exp\left(-\frac{\left\|\mathbf{x}_i - \mathbf{x}_j\right\|^2}{\sigma}\right)$. Although, the final results will be affected by the σ in the similarity calculation. Inspired by a recent study [43], we set σ as the average of squared distances between all pairs. Meanwhile, $\boldsymbol{\mathcal{O}}\left(\mathbf{x}_i\right) = \left\{sim(\mathbf{x}_i, \mathbf{x}_j) > \boldsymbol{H}\left(\mathbf{x}_i\right)\right\}$, in which $\boldsymbol{H}\left(\mathbf{x}_i\right)$ is the average similarity of $\mathbf{x}_i$ with all the other instances. More specifically, $\boldsymbol{H}\left(\mathbf{x}_i\right) = \frac{1}{n}\sum_{j=1}^{n} sim(\mathbf{x}_i, \mathbf{x}_j)$. Therefore, only instance whose similarity information with each other is greater than the average similarity belongs to sets $\boldsymbol{\mathcal{O}}(\mathbf{x}_j)$. Such a mechanism can be regarded as the adaptive search of $\mathbf{x}_i$'s neighbors in accordance with the local density and similarity between instances within the feature space. Note that the weight of each edge is not very discriminating, we set θ to control the value of the sample's weight. If $\theta = 0$, w_{ij} is 0–1 weighting, and Gaussian kernel weighting otherwise. Due to the value of w_{ij} easy to enhance the optimization complexity, and unable to reach an infinity. Generally speaking, the value of parameter θ is estimated by cross-validation that is put into practice in a part of the data set.

3.2.2. Semi-supervised learning with consistency assumption

In previous studies of semi-supervised multi-label learning, the consistency assumption is considered as an effective strategy to improve the classifier's performance. To be specific, the labeled data are used to train a classifier, while the unlabeled data are used to improve the classifier by enforcing consistency constraints. The consistency constraints ensure that the classifier predictions on the unlabeled data are consistent with its predictions on other similar unlabeled data points. Thus, we estimate the label information of unlabeled data with consistency assumption. We construct the propagation matrix $\boldsymbol{P} = \boldsymbol{D}^{-1/2}\boldsymbol{W}\boldsymbol{D}^{-1/2}$ by normalizing the columns of $\boldsymbol{W}$, in which $\boldsymbol{D}$ is a diagonal matrix with $d_i = \sum_{j=1}^{n} w_{ij}$. Thus, we can define the optimization framework as follow:

$$\mathcal{L}_{con}(\boldsymbol{F}, \boldsymbol{Z}) = \frac{1}{2}\left(\sum_{i,j=1}^{n} w_{ij}\left\|\frac{f_i}{\sqrt{d_{ii}}} - \frac{f_j}{\sqrt{d_{jj}}}\right\|^2\right) + \frac{\mu}{2}\sum_{i=1}^{l}\left\|f_i^l - z_i\right\|^2, \tag{10}$$

where $\boldsymbol{F} \in \mathbb{R}^{n\times q}$ is the optimal prediction labels for all the instances $\boldsymbol{X}$, and $\boldsymbol{Z} \in \mathbb{R}^{l\times q}$ is a fitting label space to the labeled data. Moreover, f_i and z_i are the ith row vector of $\boldsymbol{F}$ and $\boldsymbol{Z}$ respectively, f_i^l is the ith row vector of $\boldsymbol{F}$ from labeled data, μ is the hyper-parameter that balances the two terms. Essentially, the first term of Eq. (10) corresponds to the smoothness constraint, guaranteeing that the solution divides similar scores to samples that are connected by large weights. Meanwhile, the second term of Eq. (10) denotes the fitting constraint, which guarantees that the solution is near to the original label assignment.

3.3. The objective function of SMCLP

As discussed above, we build the final learning framework for imbalanced multi-label learning. To be specific, we present the objective function in the following formulation:

$$\begin{aligned}\mathcal{L}(\boldsymbol{F}, \boldsymbol{Z}) =& \frac{1}{2}\left(\sum_{i,j=1}^{n} w_{ij}\left\|\frac{f_i}{\sqrt{d_{ii}}} - \frac{f_j}{\sqrt{d_{jj}}}\right\|^2\right) + \frac{\mu}{2}\left\|\boldsymbol{F}^l - \boldsymbol{Z}\right\|_{\mathcal{F}}^2 \\ &+ \frac{\lambda}{2}\|\boldsymbol{B}\boldsymbol{Z} - \boldsymbol{Y}\boldsymbol{K}\|_{\mathcal{F}}^2 + \frac{\zeta}{2}\|\boldsymbol{Z}\boldsymbol{Q} - \boldsymbol{Y}\boldsymbol{K}\|_{\mathcal{F}}^2,\end{aligned} \tag{11}$$

where $\boldsymbol{F}^l$ is the prediction label matrix of labeled instance, $\boldsymbol{B} = (1-\alpha)\boldsymbol{I} + \alpha\boldsymbol{C}$, $\boldsymbol{Q} = (1-\alpha)\boldsymbol{I} + \alpha\boldsymbol{S}$, $\boldsymbol{I}$ denotes the identity matrix, μ, λ and ζ are three trade-off parameters. It can be observed in Eq. (11) that the optimal prediction label matrix $\boldsymbol{F}$ and the fitting label space $\boldsymbol{Z}$ are included. This inclusion causes the prediction results of unlabeled instances to be influenced by both the consistency assumption and the collaborative approach. By optimizing Eq. (11), the optimal prediction label matrix $\boldsymbol{F}$ for all instances $\boldsymbol{X}$ can be obtained, thus achieving imbalanced multi-label learning.

3.4. Optimization

Notice that the optimization objective involves two independent variables. One of the variables is the optimal prediction label matrix $\boldsymbol{F}$, and the other variable is the fitting label space $\boldsymbol{Z}$. Nevertheless, it would be quite difficult to solve the two variables jointly. Thus, an alternating method is adopted to give a solution to this optimization problem. Then, we initialize $\boldsymbol{F}^{(0)} = \begin{bmatrix}\boldsymbol{Y}\\ \boldsymbol{0}\end{bmatrix}$ and $\boldsymbol{Z}^{(0)} = \boldsymbol{Y}$.

3.4.1. Update F

We first solve the $\boldsymbol{F}$ in Eq. (11) while the intermediate variable $\boldsymbol{Z}$ is fixed. Accordingly, we convert Eq. (11) into the following one:

$$\Xi(\boldsymbol{F}) = \frac{1}{2}\left(\sum_{i,j=1}^{n} w_{ij}\left\|\frac{f_i}{\sqrt{d_{ii}}} - \frac{f_j}{\sqrt{d_{jj}}}\right\|^2\right) + \frac{\mu}{2}\left\|\boldsymbol{F}^l - \boldsymbol{Z}\right\|_{\mathcal{F}}^2, \tag{12}$$

The Eq. (12) can be regarded as a soft label propagation algorithm with intermediate variable $\boldsymbol{Z}$. Thus, we can achieve $\boldsymbol{F}$ by letting the partial derivatives as zero. With some algebraic, we have:

$$\frac{\partial \Xi(\boldsymbol{F})}{\partial \boldsymbol{F}} = \boldsymbol{F} - \boldsymbol{P}\boldsymbol{F} + \mu\left(\boldsymbol{F} - \tilde{\boldsymbol{Z}}^{(t)}\right), \tag{13}$$

where $\boldsymbol{P} \in \mathbb{R}^{l\times l}$ is the propagation matrix, $\tilde{\boldsymbol{Z}}^{(t)} = \begin{bmatrix}\boldsymbol{Z}^{(t)}\\ \boldsymbol{0}\end{bmatrix}$. Next, we represent a gradient descent step with learning rate β for $\boldsymbol{F}$ such that,

$$\begin{aligned}\boldsymbol{F}^{(t+1)} &= \boldsymbol{F}^{(t)} - \beta\left(\boldsymbol{F}^{(t)} - \boldsymbol{P}\boldsymbol{F}^{(t)} + \mu\left(\boldsymbol{F}^{(t)} - \tilde{\boldsymbol{Z}}^{(t)}\right)\right) \\ &= ((1-\beta-\beta\mu)\boldsymbol{I} + \beta\boldsymbol{P})\boldsymbol{F}^{(t)} + \beta\mu\tilde{\boldsymbol{Z}}^{(t)}.\end{aligned} \tag{14}$$

3.4.2. Update Z

The fitting label space $\boldsymbol{Z}$ can be solved while the optimal prediction label matrix $\boldsymbol{F}$ is fixed. Thus, the original optimization task in Eq. (11) is reduced as follow:

$$\Psi(\boldsymbol{Z}) = \frac{\mu}{2}\left\|\boldsymbol{F}^l - \boldsymbol{Z}\right\|_{\mathcal{F}}^2 + \frac{\lambda}{2}\|\boldsymbol{B}\boldsymbol{Z} - \boldsymbol{Y}\boldsymbol{K}\|_{\mathcal{F}}^2 + \frac{\zeta}{2}\|\boldsymbol{Z}\boldsymbol{Q} - \boldsymbol{Y}\boldsymbol{K}\|_{\mathcal{F}}^2, \tag{15}$$

Placing the partial derivative of the objective function in Eq. (15) with respect to $\boldsymbol{Z}$ as 0, we then acquire the closed-form solution:

$$\gamma\boldsymbol{B}^T\boldsymbol{B}\boldsymbol{Z} + \boldsymbol{Z}(\boldsymbol{I} + \eta\boldsymbol{Q}\boldsymbol{Q}^T) = \boldsymbol{F}^l + \eta\boldsymbol{Y}\boldsymbol{K}\boldsymbol{Q}^T - \zeta\boldsymbol{B}^T\boldsymbol{Y}\boldsymbol{K} \tag{16}$$

where $\eta = \frac{\lambda}{\mu}$. Sylvester equation is very familiar so it will be efficiently solved using the Bartels-Stewart algorithm [50]. Here, the lyap[1] function is utilized to solve the mathematical issue in Matlab.

3.5. Predictive model induction

When the optimized solution of $\boldsymbol{F}$ and $\boldsymbol{Z}$ are found out, we discretize the predictive label matrix of unlabeled data. In accordance with the constraint conditions of the label information, we employ a function $g(\cdot)$ to normalize and binarize the results. Thus, the final label $\hat{\boldsymbol{Y}}^u$ of unlabeled data in data set $\boldsymbol{X}$ is predicted as follow:

$$\hat{\boldsymbol{Y}}^u = g\left(\boldsymbol{F}^{\mathrm{u}}\boldsymbol{Q}\right), \tag{17}$$

where $\boldsymbol{F}^{\mathrm{u}}$ is the optimal prediction label matrix of unlabeled data.

The pseudo code of our proposed method is demonstrated in Algorithm 1. Firstly, we calculate the correlation matrices $\boldsymbol{S}$ and $\boldsymbol{C}$. Next, we compute the label regularization matrix $\boldsymbol{K}$. After that, we use the weighted graph to capture the most discriminative data characteristics information. Then, with more information extracted from correlations between labels and correlations between instances, the class-imbalance problem in the semi-supervised multi-label data set can be effectively reduced. At last, we utilize label propagation to propagate the discriminant information and obtain the final label of

[1] https://ww2.mathworks.cn/help/control/ref/lyap.html.

the unlabeled instance. On account of the proposed formulation being convex regarding each variable, this alternating optimization process is guaranteed to converge [49].

Algorithm 1 The pseudo-code of SMCLP.

Input: Train data set $\boldsymbol{X}$, the target of labeled data $\boldsymbol{Y}$, parameters θ, α, γ, δ, τ, μ, λ, ζ and β.
Output: Final prediction $\hat{\boldsymbol{Y}}$.
Process:
1: **Initialization:** $\boldsymbol{F}^{(0)} = \begin{bmatrix} \boldsymbol{Y} \\ \boldsymbol{0} \end{bmatrix}$, $\boldsymbol{Z}^{(0)} = \boldsymbol{Y}$ and $t = 0$;
2: Construct the correlation matrices $\boldsymbol{S}$ and $\boldsymbol{C}$;
3: Compute the label regularization matrix $\boldsymbol{K}$ using Eq. (3);
4: Compute the affinity matrix $\boldsymbol{W}$ according to Eq. (9);
5: Build the undirected graph $\boldsymbol{G} = \langle \boldsymbol{E}, \boldsymbol{V}, \boldsymbol{W} \rangle$;
6: **While** not converging **do**
7: Obtain the optimal $\boldsymbol{F}^{(t+1)}$ by using Eq. (14);
8: Calculate the optimal $\boldsymbol{Z}^{(t+1)}$ via Eq. (16);
9: $t \leftarrow t + 1$;
10: **end While**
11: Induce $\hat{\boldsymbol{Y}}^{u}$ of unlabeled instance by Eq. (17);
12: **Return** $\hat{\boldsymbol{Y}}$.

4. Experiments

4.1. Data sets

In order to substantiate the exceptional performance of the proposed method, some experiments are presented on 10 multi-label data sets from Mulan Library.[2] The characteristics of these data sets are described in Table 2, in which #Instance, #Feature, #Label, and Type denote the amount of instances, amount of features, amount of labels, and feature types, respectively. Besides, Lcard stands for the average number of labels per instance, Lden represents the normalized Lcard by the total amount of class labels, DL denotes the number of distinct relevant label sets, and PDL represents normalized DL by the number of instances [10]. Besides, the Imbalance Ratio is utilized to discover the complexity of the imbalance issue, in which minIR and maxIR are the minimum and maximum imbalance ratios of all the labels, respectively. Moreover, avgIR= $\sum_{j=1}^{l} IR_j$, and $IR_j = \max\left(D_j^+, D_j^-\right) / \min\left(D_j^+, D_j^-\right)$, where D_j^+ and D_j^- are the number of positive and negative instances from label j [51]. These multi-label data sets are connected with numerous real world assignments. For example, *Birds* and *CAL500* are for audio classification, *Emotions* is applied to music categorization, and *Image* is for image classification. Four data sets are utilized to text categorization, which includes *Enron*, *Medical*, *Science*, and *Tmc2007*. In addition, two data sets are in the biology domain, in which *Genbase* is the data set for the protein function classification, and *Yeast* is applied to gene function classification. These data sets have diversified multi-label properties which provided a strong foundation for comprehensive comparative research. Furthermore, we randomly partition each data set into 10 uniform folds, with one fold being held out for testing at a time, while the remaining data is combined for training.

4.2. Evaluation metrics

In this study, we utilize four frequently used multi-label evaluation metrics, that is, Area Under Receiver Operating Characteristic Curve (AUC), Ranking Loss (RL), Average Precision (AP), and One Error (OE) to evaluate the performance. Let $\boldsymbol{D_{te}} = \left\{ \left(x_1, y_1\right), \left(x_2, y_2\right), \ldots, \left(x_{n_{te}}, y_{n_{te}}\right) \right\}$ be a test data set, where n_{te} represents the amount of test instances. Moreover, $\tilde{y}_i$ represents the label prediction vector of x_i. Suppose $rank_f(x_i, y)$ denotes the rank of ground truth label y in the predicted label ranking, which is sorted in descending order. These evaluation metrics are defined as follows,

- **AUC** [52]: The metric assesses the variation between true positive (TP) and false negative (FN) of all labels. More specifically, we require ranking all classes in the descending order of their scores, and then changing the amount of predicted class labels from 1 to q and receiving the ROC curve through calculating TP rate and FP rate for each predicted class labels. At last, we calculate the area under the ROC curve as the final result.
- **RL** [53]: The metric assesses the fraction that an irrelevant label is ranked above a relevant label. Suppose $\boldsymbol{Y}_i$ and $\overline{\boldsymbol{Y}_i}$ are the sets of relevant and irrelevant labels related with the test instance x_i.

$$RL = \frac{1}{n_{te}} \sum_{i=1}^{n_{te}} \frac{1}{\left|Y_i\right| \left|\overline{Y_i}\right|} \{(y_j, y_k) | rank_f\left(x_i, y_j\right) \geq rank_f(x_i, y_k), \left(y_j, y_k\right) \in \left|\boldsymbol{Y}_i\right| \times \left|\overline{\boldsymbol{Y}_i}\right| \}. \tag{18}$$

- **AP** [54]: The metric assesses the average fraction of relevant labels ranked above a specific relevant label c.

$$AP = \frac{1}{n_{te}} \sum_{i=1}^{n_{te}} \frac{1}{|\boldsymbol{Y}|} \sum_{y \in \boldsymbol{Y}} \frac{\left|\left\{c | rank_f(x_i, y) \leq rank_f\left(x_i, c\right)\right\}\right|}{rank_f\left(x_i, y\right)}. \tag{19}$$

- **OE** [54]: The metric assesses how many times the top-ranked label is not in the set of relevant labels of the instance.

$$OE = \frac{1}{n_i} \sum_{i=1}^{n_i} \left| \min_{y \in Y_i} rank_f\left(x_i, y\right) \notin \boldsymbol{Y} \right|. \tag{20}$$

These metrics evaluate the performance of multi-label algorithms in various sides. The AUC and AP are directly proportional to the algorithm's efficacy, with higher values indicating better performance. On the other hand, the RL and OE are inversely proportional to the algorithm's effectiveness, with lower values suggesting better performance [35].

4.3. Experimental settings

To verify the superiority of the proposed method, we select seven representative methods as comparing methods. A brief introduction of these methods is given as follows:

- **CMLP** [47]: Promising semi-supervised multi-label learning method through involving collaboration technique to exploit label correlations, which frequently outperforms other existing semi-supervised multi-label methods.
- **TRAM** [44]: Superior semi-supervised multi-label method with a label set propagation, which formulates the problem via utilizing hard label method to detect communities on the graph.
- **GCLE** [35]: State-of-the-art imbalanced multi-label learning method, which uses label enhancement approach recovered the logical label to numerical label and trains the induction model in both single label data sets and multi-label data sets via graph constraints.
- **COCOA** [10]: Recently representative imbalanced multi-label learning method, which considers the randomness in pairwise coupling and utilizes ensemble learning strategy to combine the predictions of randomly generated imbalanced learners.
- **CPNL** [33]: Recently superior cost sen-sensitive multi-label learning method, which explores the positive and negative label pairwise correlations to solve the class-imbalance issue.
- **BR** [23]: Transformation based multi-label learning method, which transforms the original multi-label learning issue into numerous independent binary learning problem, one per class label.
- **CLR** [26]: Second-order based multi-label learning method, which leverages calibration labels to rank all relevant and irrelevant labels.

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

Table 2
Characterization of multi-label data sets used in this study.

Data sets	#Instance	#Feature	#Label	Lcard	Lden	DL	PDL	Imbalance ratio			Type	Domain
								minIR	maxIR	avgIR		
Birds	645	260	20	1.470	0.073	133	0.206	1.194	106.500	31.276	Numeric	Audio
CAL500	502	68	174	26.044	0.150	502	1.000	1.041	99.400	22.345	Numeric	Audio
Emotions	593	72	6	1.868	0.311	27	0.046	1.247	3.007	2.320	Numeric	Music
Enron	1 702	1001	53	3.378	0.064	753	0.442	1.009	1701.000	136.867	Nominal	Text
Genbase	662	1185	27	1.252	0.046	32	0.048	2.871	661.000	143.452	Nominal	Biology
Image	2 000	294	5	1.236	0.247	20	0.010	2.448	3.890	3.116	Numeric	Images
Medical	978	1449	45	1.245	0.027	94	0.096	2.677	977.000	328.069	Numeric	Text
Science	5 000	743	40	1.451	0.036	398	0.080	3.259	1249.000	203.463	Numeric	Text
Tmc2007	28 596	500	22	2.220	0.101	1172	0.041	1.449	69.958	27.996	Nominal	Text
Yeast	2 417	103	14	4.238	0.303	198	0.082	1.329	70.088	8.950	Numeric	Biology

Above comparing methods can be divided into three series. Firstly, the proposed method is compared against two methods that are competent for tackling the semi-supervised multi-label learning issue, including CMLP [47] and TRAM [44]. Secondly, the proposed method is compared to three superior methods which are capable of solving the imbalanced multi-label learning problem, i.e., GCLE [35], COCOA [10] and CPNL [33]. Finally, the proposed method is compared with two promising multi-label learning approaches BR [23] and CLR [26]. For the proposed method, parameters α and δ are turned in $\{0.01, 0.05, 0.1, 0.2, 0.5\}$, γ, τ, λ and ζ are searched in $\{0.01, 0.1, 1, 10, 100\}$. To be prudent, we set θ as 2 to improve the efficiency. Moreover, the parameters μ and β are empirically set to 0.1 as the literature [47] suggested. For the chosen comparing methods, the parameter of each approach is set based on the corresponding reference suggested. Moreover, we search the parameters for each approach, and report the best result on test data. To evaluate the superiority of the proposed method in a semi-supervised setting, we gradually increase the amount of labeled data by changing the labeling ratio from 5% to 35% with a step-size of 10%. Under each labeling ratio, we run a method 10 times on each data set, and obtain distinct results and calculate the average result for comparison analysis.

4.4. Experimental results

We demonstrate the result of each method on four evaluation metrics, as shown in Tables 3–6. In four tables, the best performance among all the approaches on each data set is highlighted in boldface. According to these results from Tables 3–6, we have numerous observations:

(a) In terms of all evaluation metrics, the proposed method can achieve comparable or superior performance with respect to most cases. Most importantly, the proposed method ranks 1st on each metric when the labeling ratio is 5%. This suggests that the proposed method is robust with a limited number of training instances.
(b) The proposed method outperforms two superior semi-supervised multi-label learning algorithms in more than 91.25% cases in regard to four evaluation metrics. For instance, apart from the *Enron* and *Tmc2007* data sets, the proposed method gets the best performance on each evaluation metric when the labeling ratio is 15%. A similar phenomenon can be witnessed with respect to labeling ratios ranging from 25% to 35%. Thus, we can conclude that the proposed method is effective for semi-supervised multi-label learning and has the advantage compared with other methods.
(c) The proposed method is superior to state-of-art imbalanced multi-label learning methods. Specially, compared with GCLE, COCOA, and CPNL, the proposed method can outperform them on almost of multi-label data sets. However, on the *Tmc2007* data set, the proposed method is inferior to COCOA. Thus, we draw the conclusion that the proposed method has the advantage on imbalanced multi-label learning.
(d) The proposed method obtains highly comparable performance to the well-established multi-label learning methods BR and CLR on these multi-label data sets with regard to all the evaluation metrics. Moreover, the imbalanced multi-label learning methods GCLE, COCOA, and CPNL also outperform the two approaches in most cases. This indicates that it is crucial to settle the imbalanced problem for multi-label learning.

In consequence, we come to the conclusion that the proposed label propagation strategy has demonstrated potential for improving the performance of imbalanced multi-label learning. Furthermore, we perform Friedman test [55] to evaluate the statistical significance of the method comparison. For υ comparing methods and N data sets, let r_i^j be the rank of jth methods on the ith data set. Under the null-hypothesis, the Friedman statistic F_F obeys a Fisher distribution with $(\upsilon - 1)$ and $(\upsilon - 1)(N - 1)$ degrees of freedom

$$F_F = \frac{(N-1)\chi_F^2}{N(\upsilon-1)-\chi_F^2}, \tag{21}$$

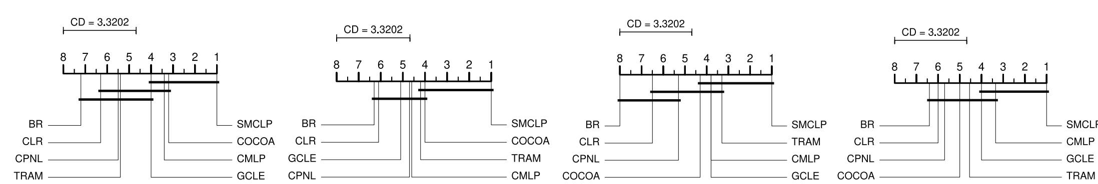
where $\chi_F^2 = \frac{12N}{\upsilon(\upsilon+1)}\left[\sum_j r_j^2 - \frac{\upsilon(\upsilon+1)^2}{4}\right]$. To reject the null hypothesis, the value of F_F must be equal to or greater than the tabled critical value at the predetermined level of significance. This critical value can be obtained from the table of F distribution in the statistical book [56]. Table 7 demonstrates the Friedman statistic F_F and the corresponding critical value of each metric. From Table 7, we can observe that all the Friedman statistics F_F are greater than the critical value, and the null hypothesis of equal performance is rejected at 0.05 significance level. Thus, we utilize the Nemenyi test to further decide which comparing methods perform statistically differently. We report the results of our proposal against the other comparing approaches when the labeling ratio is 5% in Fig. 3. According to the critical difference (CD) diagram in Fig. 3, we can safely draw the conclusion that our proposal ranks 1st in terms of all the methods when the labeling ratio is 5%. Meanwhile, we can notice that the proposed method significantly outperforms BR, CLR and CPNL, and acquires better performance than COCOA with respect to the metrics of OE, and outperforms GCLE on the metrics of RL, and while on the metrics of AUC, our proposal performs better than TRAM.

4.5. Ablation study

To validate the effectiveness of the proposed method, we conduct the following ablation experiments. To show the impact of the weighted graph construction, the proposed approach is first compared with SMCLP-1, which is regarded as a degenerated version of the proposed approach without considering the weighted graph construction. In order to reveal the influence of feature collaboration, we design a variant of the proposed method called SCMLP-2, which explores the most discriminative data characteristics by constructing the weighted graph and involves label collaboration to achieve the label of unlabeled instances. For revealing the influence of label collaboration, we also design a variant of the proposed method called SCMLP-3, which is without considering the label collaboration and utilizes the feature

Table 3
AUC of each comparing algorithm on different label ratios.

Datasets	Labeling ratio	Methods							
		SMCLP	CMLP	TRAM	GCLE	COCOA	CPNL	BR	CLR
Birds	5%	**0.7350**	0.7176	0.6815	0.7129	0.7241	0.7274	0.6354	0.6996
CAL500		**0.5383**	0.5313	0.5172	0.5099	0.5315	0.5115	0.5126	0.5095
Emotions		**0.7775**	0.7648	0.6712	0.7439	0.7488	0.7646	0.6635	0.7009
Enron		**0.6570**	0.6432	0.6221	0.6523	0.6150	0.6514	0.5541	0.5688
Genbase		**0.9301**	0.8623	0.9226	0.9196	0.9283	0.6295	0.6091	0.6421
Image		**0.7800**	0.7741	0.6976	0.7478	0.7353	0.7136	0.6765	0.7210
Medical		**0.8797**	0.8344	0.8075	0.8147	0.8615	0.8196	0.7752	0.8018
Science		**0.7029**	0.6706	0.6587	0.6913	0.6432	0.5090	0.6070	0.6257
Tmc2007		**0.7980**	0.6362	0.7074	0.7721	0.7903	0.5331	0.5531	0.5541
Yeast		**0.6114**	0.6045	0.5675	0.5955	0.6057	0.5081	0.5884	0.5873
Birds	15%	**0.7767**	0.7683	0.6918	0.7555	0.7588	0.7306	0.6883	0.7026
CAL500		**0.5476**	0.5350	0.5261	0.5386	0.5343	0.5410	0.5234	0.5162
Emotions		**0.8030**	0.7928	0.7383	0.7902	0.7933	0.7704	0.7112	0.7074
Enron		0.6701	**0.6817**	0.6263	0.6569	0.6388	0.6566	0.5553	0.5716
Genbase		**0.9594**	0.9464	0.9382	0.9539	0.9128	0.7295	0.6223	0.6423
Image		**0.8044**	0.7791	0.7159	0.7748	0.7985	0.7406	0.7076	0.7243
Medical		**0.8868**	0.8526	0.8355	0.8629	0.8701	0.8345	0.7899	0.8375
Science		**0.7454**	0.7070	0.6953	0.7295	0.6762	0.5106	0.6091	0.6260
Tmc2007		**0.8098**	0.6501	0.7339	0.8019	0.8076	0.5344	0.5536	0.5547
Yeast		**0.6318**	0.6214	0.5807	0.6188	0.6247	0.5122	0.5951	0.5938
Birds	25%	**0.8003**	0.7956	0.7202	0.7702	0.7862	0.7381	0.6901	0.7137
CAL500		**0.5534**	0.5368	0.5281	0.5404	0.5465	0.5461	0.5267	0.5165
Emotions		**0.8115**	0.7907	0.7402	0.7916	0.8035	0.8078	0.7140	0.7321
Enron		**0.6918**	0.6475	0.6290	0.6570	0.6885	0.6583	0.5670	0.5879
Genbase		**0.9615**	0.9477	0.9404	0.9560	0.9262	0.7346	0.6412	0.6439
Image		**0.8094**	0.8053	0.7229	0.7913	0.8047	0.7513	0.7087	0.7287
Medical		0.8866	0.8838	0.8873	0.8959	**0.8987**	0.8538	0.8650	0.8529
Science		**0.7602**	0.7311	0.7154	0.7469	0.6973	0.5107	0.6112	0.6269
Tmc2007		0.8128	0.6575	0.7435	0.8090	**0.8130**	0.5356	0.5550	0.5550
Yeast		**0.6421**	0.6330	0.5883	0.6317	0.6325	0.5257	0.6011	0.6039
Birds	35%	**0.8063**	0.8039	0.7362	0.7734	0.8022	0.7410	0.7127	0.7143
CAL500		**0.5671**	0.5432	0.5351	0.5410	0.5471	0.5467	0.5334	0.5259
Emotions		**0.8187**	0.8174	0.7580	0.8064	0.8154	0.8068	0.7236	0.7899
Enron		0.7066	0.6600	0.6343	**0.7072**	0.6958	0.6667	0.5752	0.5889
Genbase		**0.9662**	0.9563	0.9493	0.9601	0.9436	0.7357	0.6481	0.6397
Image		**0.8190**	0.8098	0.7264	0.7978	0.8074	0.7540	0.7362	0.7393
Medical		**0.9132**	0.9074	0.9038	0.8868	0.9025	0.8636	0.8937	0.8750
Science		**0.7711**	0.7367	0.7280	0.7590	0.7090	0.5132	0.6272	0.6426
Tmc2007		**0.8161**	0.8158	0.7493	0.8135	0.6657	0.7279	0.5568	0.5551
Yeast		**0.6490**	0.6387	0.5937	0.6412	0.6421	0.5846	0.6018	0.6069

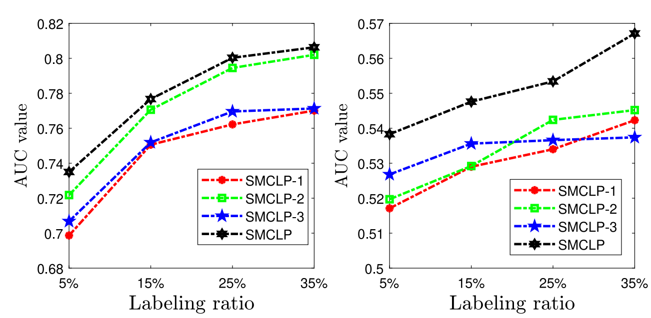


(a) AUC (b) RL (c) AP (d) OE

Fig. 3. Nemenyi test results in terms of labeling ratio as 5%.

collaboration to achieve the label of unlabeled instances. We compare the proposed approach with these variants by using AUC on the *Birds* and *CAL500* data sets while the labeling ratio varies from 5% to 30%, and the comparison results are demonstrated in Fig. 4. From Fig. 4, we list the following observations.

(a) The full fledged version of the proposed method achieves the best rank in these two data sets. Therefore, we can gain that the proposed method, which utilizes label propagation under the consistency assumption and collaboration assumption, promotes the prediction for the minority class, hence achieving imbalanced multi-label learning.
(b) Weighted graph construction based methods outperform the other compared methods in most cases. This proves that the weighted graph construction strategy can capture the most discriminative



(a) Birds (b) CAL500

Fig. 4. Ablation study of the SMCLP on Birds and CAL500 data sets.

Table 4
RL of each comparing algorithm on different label ratios.

Datasets	Labeling ratio	Methods							
		SMCLP	CMLP	TRAM	GCLE	COCOA	CPNL	BR	CLR
Birds	5%	**0.1807**	0.2295	0.2016	0.1981	0.2168	0.2017	0.2457	0.2072
CAL500		**0.2269**	0.2276	0.2310	0.2355	0.2391	0.2281	0.2382	0.2695
Emotions		**0.2417**	0.3531	0.3656	0.3217	0.2867	0.2625	0.3631	0.3002
Enron		**0.1259**	0.1396	0.1294	0.1611	0.1288	0.1297	0.1640	0.1951
Genbase		**0.0025**	0.0068	0.0103	0.0100	0.0095	0.0040	0.0035	0.0915
Image		**0.2546**	0.2822	0.2568	0.3012	0.2633	0.2840	0.3975	0.2711
Medical		**0.1104**	0.2028	0.1799	0.1471	0.1604	0.1841	0.2102	0.1553
Science		**0.1426**	0.1483	0.1431	0.1601	0.1637	0.1645	0.1795	0.1724
Tmc2007		**0.0366**	0.0667	0.0616	0.1053	0.0376	0.1650	0.0475	0.0769
Yeast		**0.2071**	0.2122	0.2182	0.2274	0.2149	0.2207	0.2214	0.2350
Birds	15%	**0.1444**	0.1618	0.1497	0.1705	0.1848	0.2013	0.2140	0.1663
CAL500		**0.1924**	0.2005	0.2187	0.2150	0.2386	0.2091	0.2369	0.2370
Emotions		**0.2040**	0.2156	0.2103	0.2287	0.2656	0.2092	0.3543	0.2166
Enron		**0.1015**	0.1223	0.1281	0.1599	0.1079	0.1191	0.1284	0.1837
Genbase		**0.0020**	0.0063	0.0070	0.0070	0.0078	0.0028	0.0031	0.0101
Image		**0.2070**	0.2192	0.2070	0.2546	0.2626	0.2657	0.3447	0.2239
Medical		0.0871	0.1239	0.0921	**0.0623**	0.1341	0.1756	0.1631	0.1015
Science		**0.1122**	0.1209	0.1382	0.1203	0.1618	0.1623	0.1791	0.1639
Tmc2007		0.0374	0.0495	**0.0359**	0.0881	0.0500	0.1644	0.0465	0.0529
Yeast		**0.1880**	0.1943	0.1897	0.2142	0.2132	0.2178	0.2057	0.2047
Birds	25%	**0.1272**	0.1411	0.1315	0.1474	0.1835	0.1842	0.2043	0.1347
CAL500		**0.1898**	0.1970	0.1941	0.2100	0.2364	0.1997	0.2359	0.2333
Emotions		0.1897	0.1937	**0.1820**	0.2206	0.2541	0.1944	0.3507	0.2034
Enron		**0.0937**	0.1209	0.1261	0.1494	0.1070	0.1069	0.1262	0.1762
Genbase		**0.0017**	0.0021	0.0038	0.0019	0.0067	0.0022	0.0029	0.0072
Image		**0.1899**	0.1992	0.2043	0.2404	0.2518	0.2536	0.3327	0.2139
Medical		**0.0513**	0.0633	0.0631	0.0550	0.1035	0.1446	0.1202	0.0753
Science		**0.1063**	0.1149	0.1262	0.1103	0.1465	0.1620	0.1601	0.1487
Tmc2007		**0.0354**	0.0450	0.0357	0.0822	0.0371	0.1643	0.0416	0.0490
Yeast		**0.1832**	0.1875	0.1866	0.2075	0.2073	0.2163	0.2051	0.1953
Birds	35%	**0.1072**	0.1229	0.1188	0.1345	0.1816	0.1771	0.2042	0.1271
CAL500		**0.1894**	0.1950	0.1900	0.2076	0.2358	0.1989	0.2309	0.2305
Emotions		**0.1737**	0.1782	0.1814	0.2129	0.1902	0.1766	0.3360	0.2019
Enron		**0.0864**	0.1129	0.1249	0.1402	0.0981	0.1035	0.1140	0.1656
Genbase		**0.0012**	0.0016	0.0017	0.0016	0.0058	0.0016	0.0028	0.0023
Image		**0.1879**	0.1901	0.1946	0.2368	0.2375	0.2502	0.3168	0.2128
Medical		**0.0319**	0.0574	0.0331	0.0353	0.0737	0.1353	0.0822	0.0445
Science		0.1037	0.1083	0.1228	**0.1022**	0.1355	0.1568	0.1598	0.1414
Tmc2007		0.0346	0.0428	**0.0322**	0.0781	0.0449	0.1639	0.0401	0.0464
Yeast		**0.1772**	0.1839	0.1799	0.2006	0.2023	0.2158	0.2048	0.1853

data characteristics and affects the performance of the label propagation methods.

(c) SCMLP-2 and SCMLP-3 are inferior to the proposed approach. The experimental result strongly demonstrates the effectiveness of label collaboration and feature collaboration. Therefore, we draw the conclusion that the proposed method under the collaboration assumption can effectively incorporate both the data distribution information and label correlations into one framework, and achieve promising results on imbalanced multi-label data sets.

4.6. Visualization analysis

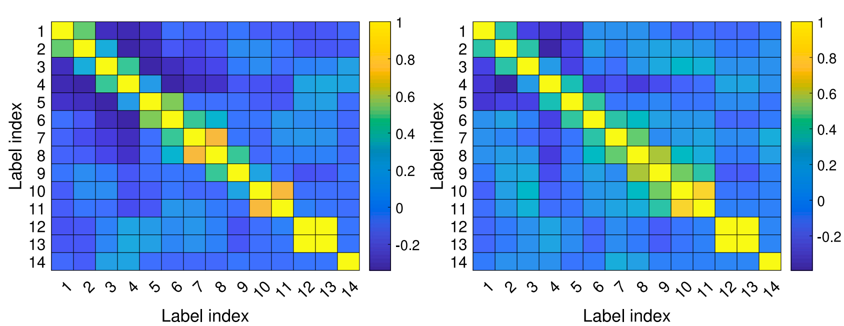
In order to illustrate the learning performance of our proposed method in terms of semi-supervised setting, a visual experiment is utilized to show the label correlations of the true label space on the unlabeled data and the label correlations of the predicted label space on the *Yeast* data set when the labeling ratio is 5%. The experimental result is demonstrated in Fig. 5, in which the horizontal axis and the vertical axis represent the label index. From Fig. 5, we can notice that the proposed method successfully detects and preserves the label correlation pattern in the ground truth data set. This further demonstrated that our proposed label propagation can efficiently achieve semi-supervised multi-label learning, and ensure the balanced propagate outcomes.

4.7. Parameter sensitivity analysis

In this study, we research the influences of the six parameters for the proposed method on four data sets, including *Birds*, *CAL500*, *Emotions*, and *Enron*. Among these parameters, α is used to reflect the collaboration degree of labels, δ is used to reflect the collaboration degree of features, γ and τ are regularization parameters to reflect the influence of the correlation matrix, and both of λ and ζ are controlling the influence of the label collaboration and feature collaboration respectively. We analyze the sensitivity of these parameters and report the results in Fig. 6. Fig. 6(a) presents the average classification results (AUC) while changing the value of α and fixing the others at their best setting. From Fig. 6(a), we can observe that the proposed method leans towards obtaining better performance with the enlarging of the value of α, and the best performance is obtained at some intermediate values. After that, when increasing α, the curves keep stable. A similar result can be witnessed for parameters δ and τ, as demonstrated in Fig. 6(b) and (d). However, we can discover that the α is more sensitive than δ and τ, and α prefers to be a smaller value than δ and τ for gaining better performance. Besides, Fig. 6(c) demonstrates the effect of parameter γ on four data sets. As is shown in the figures, with increasing γ, the performance gradually gets deteriorate. But, when γ is very large, performance starts to improve. Therefore, the proposed method is sensitive to the above four parameters. In addition, Fig. 6(e)–(f) show the influence of λ and ζ. As can be seen, the average AUC value is relatively stable while varying values of ζ and λ, and therefore the proposed method is not sensitive to the two parameters.

Table 5
AP of each comparing algorithm on different label ratios.

Datasets	Labeling ratio	Methods							
		SMCLP	CMLP	TRAM	GCLE	COCOA	CPNL	BR	CLR
Birds	5%	**0.6527**	0.6313	0.6448	0.6279	0.6062	0.5947	0.5928	0.6156
CAL500		**0.4662**	0.4339	0.4660	0.4400	0.4169	0.4564	0.3765	0.4130
Emotions		**0.7292**	0.6943	0.7196	0.6992	0.7142	0.7240	0.6427	0.6920
Enron		**0.6081**	0.5699	0.5206	0.5589	0.5551	0.4792	0.4432	0.4697
Genbase		**0.9643**	0.9589	0.9537	0.9416	0.9556	0.9291	0.9213	0.9243
Image		**0.7387**	0.7273	0.6889	0.7003	0.7057	0.6828	0.6265	0.6894
Medical		**0.7175**	0.6315	0.6780	0.6433	0.6219	0.6671	0.5447	0.5740
Science		**0.4942**	0.4525	0.4413	0.4746	0.4858	0.4159	0.3924	0.4180
Tmc2007		**0.7922**	0.6630	0.7779	0.7733	0.7448	0.5789	0.4376	0.4454
Yeast		**0.7285**	0.7143	0.7223	0.7176	0.7123	0.7071	0.6700	0.6994
Birds	15%	**0.6936**	0.6694	0.6877	0.6884	0.6100	0.6231	0.6024	0.6230
CAL500		**0.4922**	0.4836	0.4764	0.4441	0.4541	0.4646	0.3950	0.4236
Emotions		**0.7783**	0.7606	0.7611	0.7042	0.7596	0.7568	0.6458	0.7261
Enron		**0.6334**	0.6311	0.5475	0.5596	0.5615	0.4910	0.4748	0.4864
Genbase		**0.9906**	0.9805	0.9686	0.9447	0.9861	0.9796	0.9312	0.9350
Image		**0.7634**	0.7430	0.7598	0.7133	0.7444	0.6999	0.6589	0.6967
Medical		**0.7319**	0.6676	0.7210	0.7029	0.6986	0.6873	0.6201	0.6723
Science		0.5392	0.5233	0.5376	0.5366	0.4904	**0.5502**	0.4172	0.4572
Tmc2007		**0.8191**	0.7460	0.8151	0.7804	0.8153	0.5806	0.4396	0.4481
Yeast		**0.7382**	0.7333	0.7340	0.7188	0.7316	0.7139	0.6853	0.7015
Birds	25%	0.7147	0.6954	**0.7169**	0.6932	0.7085	0.6236	0.6131	0.6232
CAL500		**0.4939**	0.4854	0.4860	0.4923	0.4642	0.4739	0.4012	0.4280
Emotions		**0.7854**	0.7718	0.7819	0.7513	0.7749	0.7849	0.6979	0.7303
Enron		**0.6609**	0.6457	0.6413	0.5610	0.5752	0.5211	0.4825	0.4943
Genbase		**0.9934**	0.9864	0.9843	0.9602	0.9909	0.9873	0.9327	0.9360
Image		**0.7752**	0.7645	0.7626	0.7682	0.7593	0.7014	0.6642	0.6975
Medical		**0.7644**	0.7189	0.7364	0.7347	0.7232	0.7221	0.7033	0.7069
Science		**0.5694**	0.5424	0.5444	0.5605	0.5675	0.5540	0.4174	0.4767
Tmc2007		**0.8262**	0.7647	0.8247	0.7911	0.8224	0.5808	0.4398	0.4483
Yeast		**0.7453**	0.7437	0.7431	0.7205	0.7426	0.7208	0.6906	0.7054
Birds	35%	**0.7469**	0.7237	0.7288	0.7302	0.7372	0.6265	0.6180	0.6311
CAL500		**0.4989**	0.4957	0.4937	0.4600	0.4685	0.4889	0.4052	0.4291
Emotions		0.7870	0.7867	**0.7928**	0.7699	0.7831	0.7854	0.7739	0.7578
Enron		**0.6684**	0.6559	0.5630	0.6436	0.6006	0.5394	0.4873	0.5219
Genbase		**0.9918**	0.9874	0.9959	0.9879	0.9917	0.9851	0.9432	0.9741
Image		**0.7838**	0.7772	0.7743	0.7760	0.7623	0.7040	0.6734	0.7058
Medical		0.7876	**0.7879**	0.7777	0.7763	0.7712	0.7840	0.7295	0.7583
Science		**0.5883**	0.5595	0.5878	0.5769	0.5774	0.5639	0.4261	0.5029
Tmc2007		**0.8311**	0.7864	0.8304	0.8072	0.8272	0.7449	0.4440	0.4668
Yeast		**0.7527**	0.7461	0.7518	0.7317	0.7491	0.7233	0.7012	0.7219



(a) True label space (b) Predicted label space

Fig. 5. Visualization of multi-label correlations.

4.8. Convergence analysis

The proposed method contains two unknown variables, that is, the optimal prediction label matrix $\boldsymbol{F}$ and the fitting label space $\boldsymbol{Z}$. In the process of solving the optimization problem, we employ an iterative optimization approach to search for the optimal solution. As one of them converges, the objective value can reach stability. Fig. 7 describes the vary of the objective value on each iteration with respect to the *Birds*, *CAL500*, *Emotions* and *Enron* data sets when labeling ratio is 5%. From Fig. 7(a), we can witness a rapid decline in the value of the objective function over several iterations, followed by a stable state. This observation demonstrates the effectiveness of the proposed method in attaining the optimal solution. A similar phenomenon can be observed with respect to the *CAL500*, *Emotions* and *Enron* data sets, as shown in Fig. 7(b)–(d) respectively. Therefore, we draw the conclusion that the proposed method is reasonable and effective in achieving the optimal solution.

4.9. Discussion

Recently, imbalanced multi-label learning methods have gained increasing popularity in decreasing the difficulty of classification tasks.

Table 6
OE of each comparing algorithm on different label ratios.

Datasets	Labeling ratio	Methods							
		SMCLP	CMLP	TRAM	GCLE	COCOA	CPNL	BR	CLR
Birds	5%	**0.4046**	0.4176	0.4405	0.4047	0.5008	0.5355	0.4976	0.4502
CAL500		**0.1376**	0.2471	0.1383	0.2809	0.1782	0.2495	0.2034	0.5975
Emotions		**0.3688**	0.3829	0.4326	0.3741	0.4309	0.3918	0.4752	0.4139
Enron		**0.2498**	0.3055	0.4638	0.3722	0.4805	0.3686	0.5288	0.6296
Genbase		**0.0031**	0.0143	0.0143	0.0036	0.0049	0.0109	0.0087	0.0207
Image		**0.4058**	0.4415	0.4247	0.4653	0.4895	0.4932	0.5600	0.4784
Medical		**0.2871**	0.3409	0.5183	0.4193	0.4828	0.7269	0.5602	0.4329
Science		**0.6229**	0.6383	0.6928	0.6610	0.7118	0.7659	0.7331	0.6278
Tmc2007		**0.1352**	0.2347	0.2430	0.2436	0.2343	0.4092	0.6570	0.6509
Yeast		**0.2499**	0.2595	0.2692	0.3204	0.2817	0.2521	0.2707	0.4384
Birds	15%	**0.3515**	0.3552	0.3862	0.3715	0.4829	0.5139	0.4878	0.4007
CAL500		**0.1140**	0.1195	0.1353	0.2546	0.1509	0.1883	0.1824	0.5550
Emotions		**0.3129**	0.3386	0.3525	0.3406	0.4113	0.3161	0.4699	0.4025
Enron		**0.2335**	0.2909	0.3324	0.3688	0.4756	0.3594	0.5022	0.5912
Genbase		**0.0023**	0.0103	0.0023	0.0032	0.0045	0.0074	0.0048	0.0089
Image		**0.3627**	0.3724	0.3970	0.3841	0.4647	0.4676	0.5458	0.4459
Medical		**0.2741**	0.3386	0.4345	0.3461	0.4018	0.7241	0.4289	0.4071
Science		**0.5447**	0.5739	0.6028	0.5744	0.7065	0.7648	0.7187	0.6265
Tmc2007		0.2121	0.2069	0.2389	0.2113	**0.1221**	0.4074	0.6558	0.6504
Yeast		**0.2495**	0.2530	0.2628	0.2545	0.2791	0.2516	0.4010	0.2627
Birds	25%	**0.3368**	0.3471	0.3740	0.3368	0.4698	0.5119	0.4731	0.3574
CAL500		**0.1124**	0.1167	0.1174	0.2389	0.1488	0.1241	0.1216	0.5199
Emotions		**0.2876**	0.3033	0.3034	0.3264	0.3759	0.2950	0.4060	0.3394
Enron		0.2701	0.2874	**0.2255**	0.3531	0.4100	0.3462	0.4923	0.5466
Genbase		**0.0020**	0.0092	0.0020	0.0021	0.0043	0.0070	0.0046	0.0056
Image		**0.3500**	0.3653	0.3600	0.3740	0.4632	0.4631	0.5005	0.4292
Medical		**0.2654**	0.3268	0.3575	0.2956	0.3708	0.7233	0.3387	0.3587
Science		**0.5272**	0.5373	0.5843	0.5427	0.6800	0.7635	0.7166	0.5981
Tmc2007		0.2022	0.2006	0.2283	0.2037	**0.1160**	0.4069	0.6502	0.6492
Yeast		**0.2427**	0.2504	0.2608	0.2482	0.2773	0.2500	0.2481	0.3905
Birds	35%	**0.3024**	0.3333	0.3310	0.3095	0.4160	0.5083	0.4519	0.3214
CAL500		**0.1070**	0.1148	0.1148	0.2294	0.1447	0.1070	0.1195	0.5076
Emotions		**0.2719**	0.2979	0.2989	0.2989	0.3109	0.2921	0.4043	0.3371
Enron		**0.2113**	0.2728	0.2178	0.3216	0.4001	0.3241	0.4688	0.4842
Genbase		**0.0017**	0.0064	0.0018	0.0019	0.0041	0.0047	0.0042	0.0046
Image		0.3377	0.3413	**0.3354**	0.3708	0.4542	0.4527	0.4958	0.4247
Medical		**0.1541**	0.2625	0.2670	0.2513	0.2956	0.7201	0.2673	0.2961
Science		**0.4997**	0.5172	0.5566	0.5234	0.6632	0.7628	0.7152	0.5738
Tmc2007		0.1991	0.1954	0.1953	0.2005	**0.1082**	0.4022	0.6446	0.6477
Yeast		**0.2265**	0.2410	0.2512	0.2392	0.2686	0.2482	0.2468	0.3766

Table 7
Friedman statistics & critical value on evaluation metrics.

Evaluation metric	F_F				Critical value
	5%	15%	25%	35%	
AUC	17.5077	27.4162	19.7343	21.6818	≈2.25
RL	7.4062	8.6471	7.0102	8.2840	
AP	26.7955	27.2500	14.6398	9.3942	
OE	14.7288	14.3441	11.2899	6.8991	

Conventional imbalanced multi-label learning methods are primarily designed for fully supervised scenarios and often struggle to extend to semi-supervised learning setting. Label propagation is an effective strategy for semi-supervised multi-label learning, which improves the learning performance by utilizing the data distribution information of unlabeled data and the label information of labeled data. In addition, the issue in terms of the class-imbalance has not been investigated in existing semi-supervised multi-label classification methods, which is more severe while limited labeled data exist. The core idea of the proposed method is to access the inherent property of class-imbalance in multi-label data, and utilize label propagation to assist in the implementation of semi-supervised imbalanced multi-label classification. Different from conventional methods, the proposed method utilized a label regularization matrix to effectively tackle the imbalanced multi-label issue. Additionally, it leverages a collaborative framework to ensure balanced outcomes. Moreover, the proposed method has a good mechanism for exploiting the representation of labeled and unlabeled data by encoding the similarity between all instances in a weighted graph manner. From the experimental results, we can conclude that the proposed method significantly enhances the performance of multi-label learning.

The proposed method also has some shortcomings. We utilized a label regularization matrix to handle the imbalanced multi-label problem in a semi-supervised setting. However, the results of computing the label regularization matrix rely on the number of instances per label. Thus, when labels never appear in the training data sets, the difficulties of classification will increase.

5. Conclusion

In this study, we proposed a label propagation strategy to deal with the imbalanced multi-label problem in a semi-supervised setting. The label correlations and data distribution information were fully utilized to construct the loss function in a collaborative manner. Meanwhile, a label regularization matrix was learned to settle the imbalanced multi-label problem. Subsequently, we extend to a semi-supervised learning scenario and construct a weighted representation of both labeled and unlabeled data by leveraging their similar properties. Then, we incorporated the consistency assumption into the loss function to enhance the semi-supervised learning performance. Finally, an efficient alternating minimization strategy was introduced to seek the optimal solution. Extensive experiments explained that the proposed method gains a better performance.

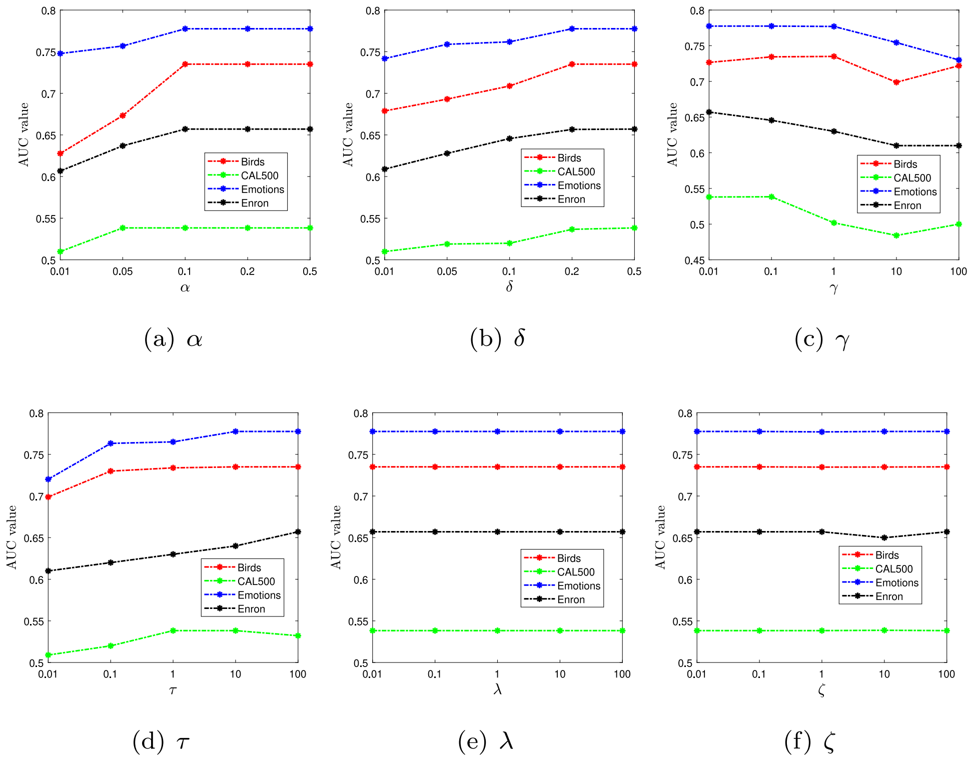


Fig. 6. Parameter analysis on *Birds*, *CAL500*, *Emotions*, and *Enron* data sets.

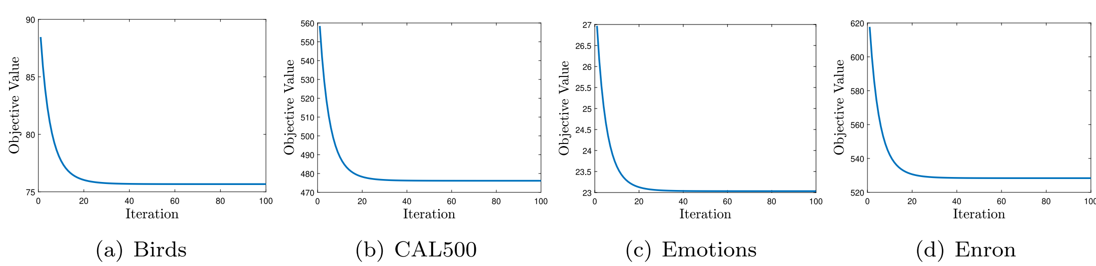


Fig. 7. Objective value of the proposed method.

It is important to note that our proposed method can handle the semi-supervised imbalanced multi-label problem. This suggests a promising direction to leverage data distribution information of unlabeled data and the label information of labeled data for semi-supervised few-shot (or zero-shot) multi-label learning [57].

CRediT authorship contribution statement

Guodong Du: Methodology, Validation, Writing – original draft, Writing – review & editing. **Jia Zhang:** Conceptualization, Funding acquisition, Methodology, Validation, Writing – original draft, Writing – review & editing, Supervision. **Ning Zhang:** Validation, Visualization, Writing – original draft, Writing – review & editing. **Hanrui Wu:** Validation, Writing – original draft, Writing – review & editing. **Peiliang Wu:** Validation, Visualization, Writing – original draft, Writing – review & editing. **Shaozi Li:** Methodology, Supervision, Validation, Writing – original draft, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (No. 62106084, 62206111, 62376232, U20A20167, 62276028), the National Natural Science Foundation of Guangdong, China (No. 2022A1515010468), the Young Talent Support Project of Guangzhou Association for Science and Technology (No. QT-2023-017), the Guangzhou Basic and Applied Basic Research Foundation, China (No. 2023A04J1058), the Central Leading Local Science and

Technology Development Fund Project of Hebei Province, China (No. 236Z1811G), the Innovation Capability Improvement Plan Project of Hebei Province, China (No. 22567626H), the S&T Program of Qinhuangdao, China (No. 202301A219), the Innovation and Entrepreneurship Training Program for College Students of Yanshan University, China (No. S202310216040), the Startup Foundation for Introducing Talent of Yanshan University, China (No. 8190550).

References

[1] J. Zhang, Z. Luo, C. Li, C. Zhou, S. Li, Manifold regularized discriminative feature selection for multi-label learning, Pattern Recognit. 95 (2019) 136–150.
[2] M. Ivasic-Kos, M. Pobar, S. Ribaric, Two-tier image annotation model based on a multi-label classifier and fuzzy-knowledge representation scheme, Pattern Recognit. 52 (2016) 287–305.
[3] S. Ran, X. Li, B. Zhao, Y. Jiang, X. Yang, C. Cheng, Label correlation embedding guided network for multi-label ECG arrhythmia diagnosis, Knowl.-Based Syst. 270 (2023) 110545.
[4] Z. Huang, J. Zhang, Z. Zhu, E.Q. Wu, K.C. Tan, Identification of autistic risk candidate genes and toxic chemicals via multilabel learning, IEEE Trans. Neural Netw. Learn. Syst. 32 (9) (2021) 3971–3984.
[5] Y. Lin, Q. Hu, J. Liu, X. Zhu, X. Wu, MULFE: multi-label learning via label-specific feature space ensemble, ACM Trans. Knowl. Discov. Data 16 (1) (2022) 1–24.
[6] N. Zheng, S. Du, J. Wang, H. Zhang, W. Cui, Z. Kang, T. Yang, B. Lou, Y. Chi, H. Long, M. Ma, Q. Yuan, S. Zhang, D. Zhang, F. Ye, J. Xin, Predicting COVID-19 in China using hybrid AI model, IEEE Trans. Cybern. 50 (7) (2020) 2891–2904.
[7] A. Elisseeff, J. Weston, A kernel method for multi-labelled classification, in: Proc. Advances in Neural Information Processing Systems, Vancouver, British Columbia, Canada, 2001, pp. 681–687.
[8] R. Wang, S. Kwong, X. Wang, Y. Jia, Active *k*-labelsets ensemble for multi-label classification, Pattern Recognit. 109 (107583) (2021) 1–17.
[9] A.N. Tarekegn, M. Giacobini, K. Michalak, A review of methods for imbalanced multi-label classification, Pattern Recognit. 118 (107965) (2021) 1–12.
[10] M. Zhang, Y. Li, H. Yang, X. Liu, Towards class-imbalance aware multi-label learning, IEEE Trans. Cybern. 52 (6) (2022) 4459–4471.
[11] İ. Yağ, A. Altan, Artificial intelligence-based robust hybrid algorithm design and implementation for real-time detection of plant diseases in agricultural environments, Biology 11 (12) (2022) 1732.
[12] L. Dai, J. Zhang, C. Li, C. Zhou, S. Li, Multi-label feature selection with application to TCM state identification, Concurr. Comput. Pract. Exp. 31 (23) (2019).
[13] Y.B. Özçelik, A. Altan, Overcoming nonlinear dynamics in diabetic retinopathy classification: A robust AI-based model with chaotic swarm intelligence optimization and recurrent long short-term memory, Fractal Fract. 7 (8) (2023) 598.
[14] J. Zhang, H. Wu, M. Jiang, J. Liu, S. Li, Y. Tang, J. Long, Group-preserving label-specific feature selection for multi-label learning, Expert Syst. Appl. 213 (118861) (2023) 1–15.
[15] F. Kang, R. Jin, R. Sukthankar, Correlated label propagation with application to multi-label learning, in: Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition, New York, NY, USA, 2006, pp. 1719–1726.
[16] W. Zhan, M. Zhang, Inductive semi-supervised multi-label learning with co-training, in: Proc. ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Halifax, NS, Canada, 2017, pp. 1305–1314.
[17] L. Jing, L. Yang, J. Yu, M.K. Ng, Semi-supervised low-rank mapping learning for multi-label classification, in: Proc. IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 2015, pp. 1483–1491.
[18] Q. Wang, G. Chechik, C. Sun, B. Shen, Instance-level label propagation with multi-instance learning, in: Proc. International Joint Conference on Artificial Intelligence, Melbourne, Australia, 2017, pp. 2943–2949.
[19] B. Guo, H. Tao, C. Hou, D. Yi, Semi-supervised multi-label feature learning via label enlarged discriminant analysis, Knowl. Inf. Syst. 62 (6) (2020) 2383–2417.
[20] W. Liu, H. Wang, X. Shen, I.W. Tsang, The emerging trends of multi-label learning, IEEE Trans. Pattern Anal. Mach. Intell. 44 (11) (2022) 7955–7974.
[21] M. Zhang, J. Fang, Y. Wang, Bilabel-specific features for multi-label classification, ACM Trans. Knowl. Discov. Data 16 (1) (2022) 1–23.
[22] L. Dai, J. Zhang, G. Du, C. Li, R. Wei, S. Li, Toward embedding-based multi-label feature selection with label and feature collaboration, Neural Comput. Appl. 35 (6) (2023) 4643–4665.
[23] M.R. Boutell, J. Luo, X. Shen, C.M. Brown, Learning multi-label scene classification, Pattern Recognit. 37 (9) (2004) 1757–1771.
[24] M. Zhang, Z. Zhou, ML-KNN: A lazy learning approach to multi-label learning, Pattern Recognit. 40 (7) (2007) 2038–2048.
[25] M. Zhang, Z. Zhou, Multi-label neural networks with applications to functional genomics and text categorization, IEEE Trans. Knowl. Data Eng. 18 (10) (2006) 1338–1351.
[26] J. Fürnkranz, E. Hüllermeier, E.L. Mencía, K. Brinker, Multilabel classification via calibrated label ranking, Mach. Learn. 73 (2) (2008) 133–153.
[27] J. Read, B. Pfahringer, G. Holmes, E. Frank, Classifier chains for multi-label classification, Mach. Learn. 85 (3) (2011) 333–359.
[28] G. Tsoumakas, I.P. Vlahavas, Random *k*-labelsets: An ensemble method for multilabel classification, in: Proc. European Conference on Machine Learning, Warsaw, Poland, 2007, pp. 406–417.
[29] M. Han, H. Wu, Z. Chen, M. Li, X. Zhang, A survey of multi-label classification based on supervised and semi-supervised learning, Int. J. Mach. Learn. Cybern. 14 (3) (2023) 697–724.
[30] Z. Zhang, Z. Zhang, J. Yao, L. Liu, J. Li, G. Wu, X. Wu, Multi-label feature selection via adaptive label correlation estimation, ACM Trans. Knowl. Discov. Data 17 (9) (2023) 134:1–134:28.
[31] Y. Xu, H. Ye, N. Zhang, G. Du, Leveraging autoencoder and focal loss for imbalanced data classification, in: Proc. 12th International Conference on Information Technology in Medicine and Education, Xiamen, China, 2022, pp. 502–506.
[32] F. Charte, A.J. Rivera, M.J. del Jesus, F. Herrera, MLSMOTE: Approaching imbalanced multilabel learning through synthetic instance generation, Knowl.-Based Syst. 89 (2015) 385–397.
[33] G. Wu, Y. Tian, D. Liu, Cost-sensitive multi-label learning with positive and negative label pairwise correlations, Neural Netw. 108 (2018) 411–423.
[34] B. Wu, S. Lyu, B. Ghanem, Constrained submodular minimization for missing labels and class imbalance in multi-label learning, in: Proc. AAAI Conference on Artificial Intelligence, Phoenix, Arizona, USA, 2016, pp. 2229–2236.
[35] G. Du, J. Zhang, M. Jiang, J. Long, Y. Lin, S. Li, K.C. Tan, Graph-based class-imbalance learning with label enhancement, IEEE Trans. Neural Netw. Learn. Syst. 34 (9) (2023) 6081–6095.
[36] R. Wang, S. Chen, Y. Yu, Extending version-space theory to multi-label active learning with imbalanced data, Pattern Recognit. 142 (2023) 109690.
[37] M. Han, H. Zhang, Multiple kernel learning for label relation and class imbalance in multi-label learning, Inform. Sci. 613 (2022) 344–356.
[38] A. Braytee, W. Liu, A. Anaissi, P.J. Kennedy, Correlated multi-label classification with incomplete label space and class imbalance, ACM Trans. Intell. Syst. Technol. 10 (5) (2019) 1–26.
[39] F. Zhao, Y. Guo, Semi-supervised multi-label learning with incomplete labels, in: Proc. International Joint Conference on Artificial Intelligence, Buenos Aires, Argentina, 2015, pp. 4062–4068.
[40] A. Tan, J. Liang, W. Wu, J. Zhang, Semi-supervised partial multi-label classification via consistency learning, Pattern Recognit. 131 (108839) (2022) 1–14.
[41] J. Zhang, S. Li, M. Jiang, K.C. Tan, Learning from weakly labeled data based on manifold regularized sparse model, IEEE Trans. Cybern. 52 (5) (2022) 3841–3854.
[42] L. Wu, M. Zhang, Multi-label classification with unlabeled data: An inductive approach, in: Proc. Asian Conference on Machine Learning, Canberra, ACT, Australia, 2013, pp. 197–212.
[43] B. Raducanu, F. Dornaika, A supervised non-linear dimensionality reduction approach for manifold learning, Pattern Recognit. 45 (6) (2012) 2432–2444.
[44] X. Kong, M.K. Ng, Z. Zhou, Transductive multilabel learning via label set propagation, IEEE Trans. Knowl. Data Eng. 25 (3) (2013) 704–719.
[45] Q. Zhang, G. Zhong, J. Dong, A graph-based semi-supervised multi-label learning method based on label correlation consistency, Cogn. Comput. 13 (6) (2021) 1564–1573.
[46] S. Hu, D. Miao, W. Pedrycz, Multi granularity based label propagation with active learning for semi-supervised classification, Expert Syst. Appl. 192 (2022) 116276.
[47] H. Wang, Z. Li, J. Huang, P. Hui, W. Liu, T. Hu, G. Chen, Collaboration based multi-label propagation for fraud detection, in: Proc. International Joint Conference on Artificial Intelligence, 2020, pp. 2477–2483.
[48] H. Dong, Y. Li, Z. Zhou, Learning from semi-supervised weak-label data, in: Proc. AAAI Conference on Artificial Intelligence, New Orleans, Louisiana, USA, 2018, pp. 2926–2933.
[49] L. Feng, B. An, S. He, Collaboration based multi-label learning, in: Proc. AAAI Conference on Artificial Intelligence, Honolulu, Hawaii, USA, 2019, pp. 3550–3557.
[50] E. Kodirov, T. Xiang, S. Gong, Semantic autoencoder for zero-shot learning, in: Proc. IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 2017, pp. 4447–4456.
[51] X. Jing, X. Zhang, X. Zhu, F. Wu, X. You, Y. Gao, S. Shan, J. Yang, Multiset feature learning for highly imbalanced data classification, IEEE Trans. Pattern Anal. Mach. Intell. 43 (1) (2021) 139–156.
[52] T. Fawcett, An introduction to ROC analysis, Pattern Recognit. Lett. 27 (8) (2006) 861–874.
[53] S. Destercke, Multilabel predictions with sets of probabilities: The Hamming and ranking loss cases, Pattern Recognit. 48 (11) (2015) 3757–3765.
[54] G. Madjarov, D. Kocev, D. Gjorgjevikj, S. Dzeroski, An extensive experimental comparison of methods for multi-label learning, Pattern Recognit. 45 (9) (2012) 3084–3104.
[55] J. Demsar, Statistical comparisons of classifiers over multiple data sets, J. Mach. Learn. Res. 7 (1) (2006) 1–30.

[56] D.J. Sheskin, Handbook of Parametric and Nonparametric Statistical Procedures, Chapman and hall/CRC, 2003.
[57] F. Zhou, X. Qi, C. Xiao, J. Wang, MetaRisk: Semi-supervised few-shot operational risk classification in banking industry, Inform. Sci. 552 (2021) 1–16.

Guodong Du received the Ph.D. degree from the Department of Artificial Intelligence, Xiamen University, Xiamen, China, in 2022. He is currently a Lecturer at the School of Information Science and Engineering, Yanshan University, China. His research interests mainly focus on data mining, machine learning and its applications in healthcare. To be more specific, he is interested in multi-label learning, data fusion, feature selection, class-imbalance learning and decision systems.

Jia Zhang received the Ph.D. degree from the Department of Artificial Intelligence, Xiamen University, Xiamen, China, in 2020. He is currently a Lecturer at the College of Information Science and Technology, Jinan University, Guangzhou, China. He is broadly interested in machine learning and data mining. He is currently working on multi-label learning, data fusion, feature selection, and weakly supervised learning.

Ning Zhang received the B.Eng. Degree from the Xiangtan University, China, in 2022. She is currently working towards the M.S. degree from the School of Information Science and Engineering, Yanshan University, China. Her research interests include machine learning, pattern recognition, and data mining.

Hanrui Wu is currently an Associate Professor with the Department of Computer Science, Jinan University, Guangzhou, China. Before that, he was a Postdoctoral Research Fellow with the Department of Mathematics, The University of Hong Kong, from 2020 to 2021. He received the B.S. and Ph.D. degrees in the School of Software Engineering from South China University of Technology, China, in 2013 and 2020, respectively. His research interests include transfer learning, hypergraph learning, recommendation systems, and brain–computer interactions.

Peiliang Wu received both the B.Sc. degree in computer science and technology from Yanshan University, China, in 2004, and the Ph.D. degree in computer application technology from Yanshan University, in 2010. He is currently a professor of artificial intelligence and intelligent robotics with the Department of computer Science and Engineering, School of Information Science and Engineering, Yanshan University. His research interests include IoT robot system, multi-agent system, artificial intelligence, robot manipulation skill learning, computer vision and pattern recognition. He was a postdoctor in The State Key Laboratory for Management and Control of Complex Systems, Institute of Automation, Chinese Academy of Sciences, China, and an academic visitor in School of Engineering and the Built Environment, Edinburgh Napier University, U.K. He has served as a Local Chair for both Artificial intelligence and Internet of Things Forum, International Future Intelligent Computing Conference 2020 and China Conference on Knowledge Graph and Semantic Computing 2022.

Shaozi Li received the B.S. degree from Hunan University, Changsha, China, in 1983, the M.S. degree from Xi'an Jiaotong University, Xi'an, China, in 1988, and the Ph.D. degree from the National University of Defense Technology, Changsha, China, in 2009. He is a Full Professor with the Department of Artificial Intelligence, Xiamen University, Xiamen, China. He has directed and completed more than 20 research projects, including several national 863 Programs, National Nature Science Foundation of China, and the Ph.D. Programs Foundation of the Ministry of Education of China. Furthermore, he has authored nearly 300 papers in journals and international conferences. His research interests include artificial intelligence and its applications, moving objects detection and recognition, machine learning, computer vision, and multimedia information retrieval. Prof. Li is the Vice Director of the Technical Committee on Collaborative Computing of China Computer Federation (CCF), and the Fujian Association of Artificial Intelligence. He is a Senior Member of ACM, IEEE and CCF.