Information Sciences 543 (2021) 454–466

Contents lists available at ScienceDirect

Information Sciences

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

Noisy label tolerance: A new perspective of Partial Multi-Label Learning

Gengyu Lyu [a], Songhe Feng [a,*], Yidong Li [a]

[a] Beijing Key Laboratory of Traffic Data Analysis and Mining, School of Computer and Information Technology, Beijing Jiaotong University, China

ARTICLE INFO

Article history:
Received 31 October 2019
Received in revised form 11 September 2020
Accepted 14 September 2020
Available online 20 September 2020

2010 MSC:
00-01
99-00

Keywords:
Partial Multi-label Learning
Non-disambiguation Strategy
Precise Label
Missing Feature
Feature Completion

ABSTRACT

Partial Multi-Label learning (PML) aims to learn from training data where each example is associated with a set of candidate labels, among which only a subset of them is correct. The major challenge of PML lies in that the training procedure is prone to be misled by the label noise. To address this problem, nearly all existing PML methods focus on solely label disambiguation, i.e., dislodging the noisy labels from the candidate label set and then utilizing the remaining credible labels for model induction. However, these remaining "credible" labels may be incorrectly identified, which thereby would have a huge adverse impact on the subsequent model induction. In this paper, in contrary to the above label disambiguation strategy, we propose a simple yet effective **N**oisy l**A**bel **T**olerated p**A**rtial multi-label **L**earning (NATAL) method, where the labeling information is considered to be precise while the feature information is assumed to be missing. Using our proposed method, the task of PML can be re-interpreted as a *Feature Completion* problem, and the desired prediction model can be directly induced from the completed feature together with all candidate labels. Extensive experimental results on various data sets clearly demonstrate the effectiveness of our proposed approach.

1. Introduction

As a novel weakly-supervised learning framework, Partial Multi-Label learning (PML) learns from the ambiguous data where each training example is associated with a candidate label set, while only a subset of them is correct [1–3]. During the training process, the correct labels of each training example are concealed in its candidate label set and not directly accessible to the learning algorithm. For example, given an image together with corresponding candidate label sets (Fig. 1), where some of the labels (*bird* and *bike*) are incorrectly annotated, PML aims to learn an accurate multi-label classifier from such ambiguous training data and assign a set of proper labels for unseen examples.

A lazy strategy to accomplish the task of learning from PML data is to treat all the candidate labels as ground-truth labels, and then employs existing multi-label learning algorithms to train the desired model [4]. However, such strategy ignores the false positive labels concealed in the candidate label set, which would significantly mislead the learning process and degrade the performance of learning model. Another straightforward strategy to learn from partial multi-label data is to apply the off-the-shelf partial label learning algorithms, where similar to PML, partial label learning also aims to learn from training

* Corresponding author.
E-mail address: shfeng@bjtu.edu.cn (S. Feng).

https://doi.org/10.1016/j.ins.2020.09.019
0020-0255/

Fig. 1. An exemplar of partial multi-label learning. In crowdsourcing image tagging, among the 7 candidate labels given by the crowsourcing annotators, only 5 of them are correct due to potential unreliable annotators.

examples with redundant labels while only one candidate label is correct. Although the excellent performance of partial label learning framework on learning from data with redundant labeling information has been demonstrated in many real-world scenarios [5,6], the strict *one instance one label* constraint still strongly limits the performance of directly applying traditional partial label learning algorithms to solve the PML problem.

Recently, some specific approaches aiming at solving the PML problem have been proposed. These approaches generally follow the disambiguation-based strategy, and can be roughly grouped into two categories: Xie et al. [1] and Sun et al. [7] focus on disambiguating the candidate labels by assigning a confidence value for each candidate label, and then optimize it in an iterative manner. Fang et al. [3] and Wang et al. [8] divide the training process into two stages, i.e., first disambiguate the candidate labels by selecting the high-confidence labels from the candidate label set, and then learn a typical multi-label learning model from the selected data. However, these two-stage disambiguation-based PML methods may suffer from the problem that the incorrectly identified "reliable" labels would significantly degrade the performance of the induced predictive model.

To solve this problem, we propose a novel **N**osiy l**A**bel **T**olerated p**A**rtial multi-label **L**earning (NATAL) method, where in contrary to existing PML methods, the redundant labels are assumed to originate from the absence of feature information (Fig. 2). Namely, the labeling information is assumed to be precise while the feature information is considered to be missing. In this way, we can consider the task of PML problem as a *Feature Completion* problem and then train the desired model from the completed feature together with all candidate labels. Specifically, we first introduce a "missing" feature matrix and incorporate it into the observed feature matrix to fit "redundant" candidate labels. Then, considering the statistical co-occurrence of redundant labels and the well-known label correlations among different labels, we constrain the "missing" feature matrix and the prediction parameter matrix to be low rank, where a dual trace norm regularization item is employed to constrain the two low rank matrics. Afterwards, both the "missing" feature matrix and the predictive model are optimized in an alternative manner by minimizing the widely-used least square loss item. Extensive experiments have demonstrated that our proposed method can achieve superior or comparable performance when compared with state-of-the-art methods.

The rest of this paper is organized as follows. Firstly, we briefly discuss the related work on multi-label learning and partial label learning, as well as partial multi-label learning. Then, we present the technical details of the proposed NATAL algorithm and describe the corresponding optimization procedure for it. Afterwards, we conduct the comparable experiments and make further experimental analysis. Finally, we conclude the whole paper.

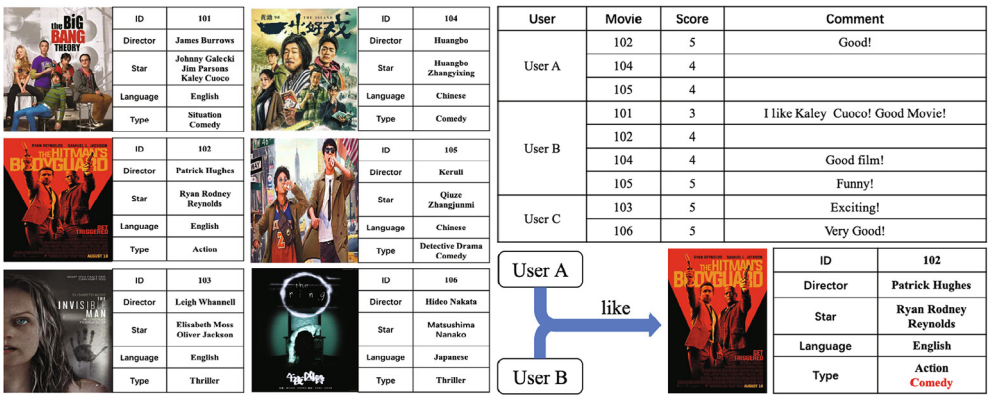

Fig. 2. An exemplar to explain why the noisy labels can be assumed to derive from missing features. In the task of movie recommendation, we aim to learn a model to achieve accurate movie recommendation for users. Among the given training data, we can observe that User A likes *Movie 102, 104, and 105* (i.e. candidate labels), and User B likes *Movie 101, 102, 104, and 105* (i.e. candidate labels). Intuitively, we can infer that both User A and User B like watching comedies. However, among these movies that they like, *Movie 102* is only recorded as an action movie instead of a comedy. Thus, we could supplement such comedy characteristic (i.e. the missing feature) for *Movie 102* (i.e. noisy label), then the redundant labels would be no longer redundant.

2. Related work

As a weakly-supervised learning framework, PML focuses on solving the multi-label learning problem with redundant labeling information, and it can be regarded as an integration of *multi-label learning* [4] and *partial label learning* [9].

2.1. Multi-Label Learning (MLL)

In multi-label learning, each example is associated with multiple valid labels simultaneously. Based on the order of label correlations being exploited for model training, existing MLL methods can be roughly characterized into three categories: *first-order strategy*, *second-order strategy* and *high-order strategy*. For the first-order strategy, MLL problem is decomposed into multiple binary classification problems [10], where the classifier for each label is trained independently. Thus, such strategy is also known as binary relevance. For the second-order strategy, pairwise label correlations are considered, where the rankings between the proper and improper labels [11], or any pair of labels [12,13] are often taken into consideration. For the high-order strategy, higher-level label correlations are considered, such as imposing all other labels' influences on each possible label [14]. Recently, some weakly supervised MLL frameworks are proposed and most of them focus on solving it with missing labeling information. Among these works, some methods treat the missing labels as negative labels and directly bring the label bias into the objective function [15,16]. Meanwhile, some methods treat missing labels as three states, i.e. positive labels +1, negative labels −1 and missing labels 0, to avoid the label bias [17]. Others treat the missing labels as latent variables and optimize them in an iterative manner [18]. In addition, there are also some methods that pursue to recover these missing labels and transfer the problem as matrix completion problem [19]. Except for the above weakly supervised MLL framework, there is also a typical weakly supervised MLL framework named Positive-Unlabeled Learning (PUL) [20,21], where some training instances are associated with positive labels and the others are not labeled. Intuitively, the unlabeled instances can be either positive or negative, since the labels are unknown. Many existing methods to deal with such problem directly regard the unlabeled instances as negative instances, and contribute them to the learning model [22].

2.2. Partial Label Learning (PLL)

In partial label learning, each example is associated with a set of candidate labels, among which only one is correct [23–26]. Existing PLL methods can be roughly grouped into three categories: *Averaging Disambiguation Strategy*, *Identification Disambiguation Strategy* and *Disambiguation-Free Strategy*. *Averaging Disambiguation Strategy*-based PLL methods usually assume that each candidate label has equal contribution to the learning model and they make prediction for unseen instances by averaging the outputs from all candidate labels [27,28]. *Identification Disambiguation Strategy*-based PLL methods often view the ground-truth label as a latent variable, and refine the model parameter in an iterative manner [29,30]. *Disambiguation-Free Strategy*-based methods adapt off-the-shelf learning techniques to learn from the partial label data, thereby they make prediction for unseen instances without any disambiguation process [31,32]. Recently, some attempts try to improve the modeling disambiguation ability by exploiting the labeling confidence of different candidate labels, where either label distribution [33,34] or label enhancement procedure [35] is often employed into its corresponding PLL framework for improving the learning model.

2.3. Partial Multi-Label learning (PML)

Partial multi-label learning learns from the ambiguous training data, where multiple labels in the candidate label set are correct [36,37]. Existing PML algorithms can be roughly divided into two categories. Some of them adopt a unified learning framework and they learn from the PML data by estimating the confidence of each candidate label, and then incorporate the estimated confidence scores into an alternative optimization procedure for model induction [1,2,7]. Others decompose the training process into two stages, where the high confidence labels are first selected from the candidate label set, and then employed for training the desired model by using some off-the-shelf MLL methods [3,8]. Whatever strategies the above algorithms follow, basically, all of them prefer to remove the noisy labels and select the high confidence labels for the PML model induction.

By contrast, in this paper, we assume that the noise labels originate from the missing feature information, and then we formulate a novel PML framework according to our proposed feature completion strategy. The details of our method are exhibited in the following section.

3. The proposed method

Formally speaking, we denote the d-dimensional feature space as $\mathcal{X} \in \mathbb{R}^d$, and the label space as $\mathcal{Y} = \{1, 2, \ldots, q\}$ with q class labels. PML aims to learn a classifier $\mathbf{f} : \mathcal{X} \mapsto 2^{\mathcal{Y}}$ from the PML training data $\mathcal{D} = \{(\mathbf{x}_i, S_i)\}(1 \leqslant i \leqslant m)$, where the instance $\mathbf{x}_i$ is described as a d-dimensional feature vector, the candidate label set $S_i \subseteq \mathcal{Y}$ is associated with the instance $\mathbf{x}_i$ and m is the number of training instances. In addition, we denote $\tilde{S}_i \subseteq S_i$ as the ground-truth label set for instance $\mathbf{x}_i$, where each $\tilde{S}_i$ corresponding to $\mathbf{x}_i$ is not directly accessible to the algorithm. Furthermore, to formulate the PML problem conveniently, we

denote $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_m]^{d\times m}$ as the training instance matrix, $\mathbf{Y} = [\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_m] \in \{0,1\}^{q\times m}$ as candidate label matrix, where $\mathbf{y}_i$ is the vector form of S_i, and $Y_{ji} = 1$ represents label j is the candidate label of instance i, while $Y_{ji} = 0$, otherwise.

3.1. Formulation

Partial multi-label learning aims to induce a multi-label model from the given PML training data and then assign accurate labels for unseen instances. To formulate such problem, a straightforward strategy is to treat all candidate labels as ground-truth ones, and then directly apply standard multi-label learning algorithm for model training. Accordingly, a general partial multi-label learning model can be learned by solving the following optimization problem **OP (1)**:

$$\min_{\mathbf{f}} L(\mathbf{X}, \mathbf{Y}; \mathbf{f}) + \Phi(\mathbf{f}) \tag{1}$$

where $L(\mathbf{X}, \mathbf{Y}; \mathbf{f})$ represents the loss function and $\Phi(\mathbf{f})$ is employed to control the model complexity.

Usually, for simplicity of the learning model, we restrict the prediction $\mathbf{f}$ to be linear functions, i.e., $f(\mathbf{X}) = \mathbf{W}^\top\mathbf{X}$, where $\mathbf{W} = [\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_q] \in \mathbb{R}^{d\times q}$ is denoted as the model parameter. Meanwhile, considering the well-known label correlations among different labels in multi-label learning, we assume that the feature mapping matrix $\mathbf{W}$ is linearly dependent to capture such label correlations, which leads $\mathbf{W}$ to be low-rank. Then, the general PML formulation can be concretized as **OP (2)**:

$$\min_{\mathbf{W}} \frac{1}{2}\left\|\mathbf{Y} - \mathbf{W}^\top\mathbf{X}\right\|_F^2 + \alpha\|\mathbf{W}\|_* \tag{2}$$

where $\|\mathbf{W}\|_*$ denotes the sum of the singular values of the matrix $\mathbf{W}$ and α is the trade-off parameter.

Obviously, the resulting of the training procedure **OP (2)** will be significantly affected by the labeling noise brought by false positive labels in the candidate label set. In this paper, in contrary to existing PML strategy that aims to remove the noisy candidate labels, we consider the label information is precise while assume the feature information is inadequate. Thus, we explicitly introduce a "missing" feature matrix $\mathbf{E}_X \in \mathbb{R}^{d\times m}$ to complete the original feature matrix $\mathbf{X}$, and then utilize the completed feature matrix to train the desired model. Afterwards, **OP (2)** can be extended as **OP (3)**:

$$\min_{\mathbf{W},\mathbf{E}_X} \frac{1}{2}\left\|\mathbf{Y} - \mathbf{W}^\top(\mathbf{X} + \mathbf{E}_X)\right\|_F^2 + \alpha\|\mathbf{W}\|_* + \lambda\|\mathbf{E}_X\|_* \tag{3}$$

where λ is the trace-off parameter, and the "missing" feature matrix $\mathbf{E}_X$ is constrained to be low-rank due to the statistical co-occurrence of redundant labels. Here, we emphasis that the "missing" features (corresponding to redundant labels) tend to be correlated since the redundant labels are correlated, especially when the scale of training examples is large enough while the number of redundant labels is not too large. Therefore, in Eq. (3), we constrain the augmented ("missing") feature $\mathbf{E}_X$ to be low-rank.

Furthermore, to address the consistency between completed feature matrix and accurate candidate label matrix, motivated by [38]394041424344, we employ a new *instance-label* consistency criterion following $\|\mathbf{Y}^\top\mathbf{Y} - (\mathbf{X} + \mathbf{E}_X)^\top(\mathbf{X} + \mathbf{E}_X)\|_F^2$, which is exploited by comparing the instance similarity based for example on visual content with the semantic similarity based on annotated labels. Accordingly, we can obtain the final optimization problem for our proposed method **OP (4)**:

$$\min_{\mathbf{W},\mathbf{E}_X} \frac{1}{2}\left\|\mathbf{Y} - \mathbf{W}^\top(\mathbf{X} + \mathbf{E}_X)\right\|_F^2 + \alpha\|\mathbf{W}\|_* + \lambda\|\mathbf{E}_X\|_* + \beta\left\|\mathbf{Y}^\top\mathbf{Y} - (\mathbf{X} + \mathbf{E}_X)^\top(\mathbf{X} + \mathbf{E}_X)\right\|_F^2 \tag{4}$$

Considering that **OP (4)** is an optimization problem with multiple variables, which is difficult to solve directly, we employ an alternating optimization procedure to update $\mathbf{W}$ and $\mathbf{E}_X$ iteratively. The details of the optimization are exhibited in the following subsection.

3.2. Optimization

Obviously, for each of the two matrices $\mathbf{W}$ and $\mathbf{E}_X$ to be solved in **OP (4)**, the objective function is convex while another matrix is kept fixed. Therefore, **OP (4)** can be solved iteratively via the following steps:

Step 1: Calculate W. Fixing the variable $\mathbf{E}_X$, we can calculate $\mathbf{W}$ by minimizing the following objective function:

$$\min_{\mathbf{W}} \frac{1}{2}\left\|\mathbf{Y} - \mathbf{W}^\top(\mathbf{X} + \mathbf{E}_X)\right\|_F^2 + \alpha\|\mathbf{W}\|_* \tag{5}$$

According to [45], Eq. (5) has the closed form solution, and the variable $\mathbf{W}$ can be updated following

$$\mathbf{W}^{t+1} = \mathcal{M}_{\alpha/\tau_{\mathbf{W}}^t}\left[\mathbf{W}^t - \frac{1}{\tau_{\mathbf{W}}^t}(\mathbf{X} + \mathbf{E}_X)\left(\mathbf{W}^{t\top}(\mathbf{X} + \mathbf{E}_X) - \mathbf{Y}\right)^\top\right]. \tag{6}$$

Here, $\tau_{\mathbf{W}}^t > \rho(\tilde{\mathbf{X}}^\top\tilde{\mathbf{X}})$ is the proximal parameter, and $\rho(\tilde{\mathbf{X}}^\top\tilde{\mathbf{X}})$ denotes the spectral radius of $\tilde{\mathbf{X}}^\top\tilde{\mathbf{X}}$, where $\tilde{\mathbf{X}} = \mathbf{X} + \mathbf{E}_X$. Moreover, $\mathcal{M}_\varepsilon[\mathbf{G}] = \mathbf{U}\mathcal{S}_\varepsilon[\mathbf{\Sigma}]\mathbf{V}^\top$, where $\mathbf{U}\mathbf{\Sigma}\mathbf{V}^\top$ is the singular value decomposition of $\mathbf{G}$, and

$$\mathcal{S}_{\varepsilon}[\mathbf{\Sigma}] = \begin{cases} \mathbf{\Sigma} - \varepsilon\ , & \text{if} \quad \mathbf{\Sigma} > \varepsilon \\ \mathbf{\Sigma} + \varepsilon, & \text{if} \quad \mathbf{\Sigma} < -\varepsilon \\ 0\ , & \text{otherwise.} \end{cases} \tag{7}$$

Algorithm 1. The Algorithm of NATAL

Input:
$\mathcal{D}$:the PML training date set $(\mathbf{x}_i, S_i)$
$\mathbf{x}^*$:the unseen instance
Parameter: α, β, λ
Output: $\mathbf{W}$, $\mathbf{E}_X$, $\mathbf{y}^*$
1: Initialize $\mathbf{W}$, $\mathbf{E}_X$ and $\alpha, \beta, \lambda, obj_{thr}$.
2: **while** $t \leqslant Iter_{max}$ **do**
3: Update $\mathbf{W}$ by solving Eq. (6).
4: Update $\mathbf{E}_X$ by solving Eq. (9).
5: Calculate the obj_{value} following **OP (4)**.
6: **if** $\|obj_{value}^{(t+1)} - obj_{value}^{(t)}\| \leqslant obj_{thr}$ **then**
7: break;
8: **end if**
9: **end while**
10: Prediction: $\mathbf{y}^* = \mathbf{W}^\top \mathbf{x}^*$.
11: **return** $\mathbf{y}^*$

Step 2: Calculate $\mathbf{E}_X$. By fixing the variable $\mathbf{W}$, the subproblem to variable $\mathbf{E}_X$ is simplified as follows:

$$\min_{\mathbf{E}_X} \frac{1}{2} \left\|\mathbf{Y} - \mathbf{W}^\top(\mathbf{X} + \mathbf{E}_X)\right\|_F^2 + \lambda \|\mathbf{E}_X\|_* + \beta \left\|\mathbf{Y}^\top\mathbf{Y} - (\mathbf{X} + \mathbf{E}_X)^\top(\mathbf{X} + \mathbf{E}_X)\right\|_F^2 \tag{8}$$

Similar to **Step 1**, Eq. (8) also has the closed form solution, and the variable $\mathbf{E}_X$ can be optimized following

$$\mathbf{E}_X^{t+1} = \mathcal{M}_{\varepsilon}\left[\mathbf{E}_X^t - \frac{\varepsilon}{\lambda}\left(\mathbf{W} \cdot \left(\mathbf{Y} - \mathbf{W}^\top(\mathbf{X} + \mathbf{E}_X^t)\right) - 4\beta(\mathbf{X} + \mathbf{E}_X^t)\left(\mathbf{Y}^\top\mathbf{Y} - (\mathbf{X} + \mathbf{E}_X^t)^\top(\mathbf{X} + \mathbf{E}_X^t)\right)\right)\right]. \tag{9}$$

where $\mathcal{M}_{\varepsilon}[\mathbf{G}] = \mathbf{U}\mathcal{S}_{\varepsilon}[\mathbf{\Sigma}]\mathbf{V}^\top$ is the singular value thresholding, $\mathcal{S}_{\varepsilon}[\mathbf{\Sigma}]$ is the shrinkage operator, $\varepsilon = \frac{2\lambda}{(1+2\beta)\tau_{\mathbf{E}_X}^t}$, $\tau_{\mathbf{W}_X} > \rho(\mathbf{W}^\top\mathbf{W})$ is the proximal parameter, and $\rho(\mathbf{W}^\top\mathbf{W})$ denotes the spectral radius of $\mathbf{W}^\top\mathbf{W}$.

During the entire optimization process, we first initialize the required variables, and then repeat the above steps until the algorithm converges or reaches the maximum iterations. Finally, we make prediction for unseen instances following $\mathbf{y}^* = \mathbf{W}^\top\mathbf{x}^*$. Algorithm 1 summarizes the whole optimization process of NATAL.

4. Experiment

4.1. Experimental setup

Since partial multi-label learning is a newly proposed learning framework [1], and there are no public PML data sets available yet. To effectively evaluate the performance of the proposed NATAL method, we implement experiments on numbers of synthetic PML data sets, where the PML data set are generated from the widely-used multi-label learning data sets under different configurations of the controlling parameter r. Here, $r \in \{1, 2, 3\}$ represents the average number of false candidate labels for training examples. Table 1 summarizes the characteristics of the experimental data sets in this paper.

Table 1
Characteristics of the experimental data sets. For each PML data set, the number of examples (EXP*), feature (FEA*), class labels (CL*), the average (AL*) and maximum (ML*) number of ground-truth labels are recorded.

Data set	Emotions	Genbase	Medical	Image	Corel5k	Bibtext	Eurlex-dc	Eurlex-sm	Configurations
EXP*	593	662	978	2000	5000	7395	19348	19348	
FEA*	72	1185	1449	294	499	1836	5000	5000	
CL*	6	27	45	5	374	159	412	201	$r \in \{1, 2, 3\}$
AL*	1.87	1.25	1.25	1.23	3.52	2.4	1.29	2.21	
ML*	3	6	3	3	5	28	7	12	

Meanwhile, we employ several baselines and some state-of-the-art methods from three categories for comparative studies, including MLL methods [**ML-KNN** [46], **RankSVM** [47]], PLL methods [**IPAL** [27], **LALO** [33]], and PML methods [**PML-fp** [1], **PARTICLE** [3], **fPML** [2], **DRAMA** [8]], where the configured parameters are utilized via the suggestions in respective literatures.

- ML-KNN [46]: A k-nearest neighbor based multi-label learning baseline that makes prediction for unseen instances via k-nearest neighbor strategy [suggested configuration: k = 10];
- RankSVM [47]: A ranking-based multi-label learning baseline that learns model by minimizing hinge loss [suggested configuration: $C \in \{10^{-3}, 10^{-2}, \ldots, 10^{3}\}$, kernel: 'RBF'];
- IPAL [27]: A k-nearest neighbor based partial label learning method that makes prediction for unseen instance by minimizing construction loss [suggested configuration: α = 0.95, k = 10];
- LALO [33]: A margin-based partial label learning method that utilizes latent label distribution to disambiguate the candidate label set [suggested configuration: $\lambda = 0.05, \mu = 0.005, k = 10$];
- PML-fp [1]: A unified partial multi-label learning method that trains the model by minimizing the ranking loss [suggested configuration: $C_3 \in \{1, 10, 100\}$]
- PARTICLE [3]: A two-stage based partial multi-label learning method [suggested configuration: $thr \in \{0.5, 0.6, \ldots, 1\}$]
- fPML [2]: An instance-based partial multi-label learning method [suggested configuration: $\lambda_2 = 1$];
- DRAMA [8]: A two-stage based partial multi-label learning method [suggested configuration: $\alpha = 0.95, thr = 0.9, k = 10$].

In addition, five popular multi-label metrics are employed to evaluate each comparing method, including *Hamming Loss*, *Ranking Loss*, *One-Error*, *Coverage* and *Average Precision*, whose detailed definitions can be found in [4] or [48].

Before conducting the experiments, we pre-introduce the values of parameters employed in our framework. Specifically, we set α, λ among $\{10^{-3}, 10^{-2}, \ldots, 10^{3}\}$, and β among $\{10^{-7}, 10^{-6}, \ldots, 10^{-3}\}$, where the three parameters are selected by ten-fold cross-validation on the training set. The initial values of τ_W and $\tau_{\mathbf{E}_X}$ are empirically set to $\|\mathbf{E}_X^{\top}\mathbf{E}_X\|_F$ and $\|\mathbf{W}^{\top}\mathbf{W}\|_F$, respectively. For each data set, ten-fold cross-validation is performed and the experimental results under different settings are separately recorded in Tables 3–5.

4.2. Experimental results

Table 3–5 separately illustrate the experimental comparisons between NATAL and other comparing methods on five evaluation metrics. Table 2 summarizes the resulting win/tie/loss counts over 24 data sets and 5 evaluation metrics. Out of 120 (8 data sets $\times$ 3 configurations $\times$ 5 metrics) statistical comparisons, the following observations can be made:

- For each comparing method, NATAL separately achieves superior or comparable performance against tailored MLL and PLL methods in 85.8% and 86.7% cases. And, it also outperforms the counterpart PML methods in 85.4% cases. In summary, the proposed NATAL algorithm is superior to state-of-the-art methods.
- For each evaluation metric, NATAL is superior or comparable to other comparing methods in 95.3% cases (*Hamming Loss*), 84.9% cases (*Ranking Loss*), 78.1% cases (*One Error*), 82.8% cases (*Coverage*) and 91.1% cases (*Average Precision*), respectively.
- Particularly, NATAL is inferior to some PLL methods on *One Error* metric, which is easy to understand, since the frameworks of PLL methods are formulated with strong *one instance one label* constraint and they only focus on the highest-level confidence label for each instance.

In order to comprehensively evaluate the superiority of the proposed NATAL, *Friedman test* [49] is utilized as the statistical test to analyze the relative performance among the comparing algorithms. According to Table 6, the null hypothesis of distinguishable performance among the comparing algorithms is rejected at 0.05 significance level. Therefore, we further employ the post hoc Bonferroni-Dunn test [49] to show the relative performance among the comparing algorithms. Fig. 3 illustrates the CD diagrams on each evaluation metric, where the average rank of each comparing algorithm is marked along the axis. According to Fig. 3, it is observed that NATAL ranks 1st on four evaluation metrics and it performs significantly superiority against half of comparing methods.

Table 2
Win/tie/loss counts of NATAL's performance against comparing methods (pairwise t-test at 0.05 significance level).

Data set	Emotions	Genbase	Medical	Image	Corel5k	Bibtext	Eurlex-dc	Eurlex-sm	Sum
Hamming Loss	22/1/1	11/13/0	11/13/0	17/1/6	13/11/0	14/10/0	12/10/2	12/12/0	112/71/9
Ranking Loss	22/2/0	6/17/1	23/1/0	7/5/12	6/3/15	24/0/0	19/5/0	17/6/1	124/39/29
One Error	17/4/3	10/10/4	21/1/2	6/2/16	19/5/0	24/0/0	17/0/7	12/2/10	126/24/42
Coverage	23/1/0	14/9/1	23/1/0	9/4/11	6/0/18	24/0/0	24/0/0	19/2/3	142/17/33
Average Precision	19/3/2	16/8/0	22/2/0	5/5/14	20/4/0	24/0/0	24/0/0	21/2/1	151/24/17
Sum	103/11/6	57/57/6	100/18/2	44/17/59	64/23/33	110/10/0	96/15/9	81/24/15	655/175/130

Table 3
Comparison of NATAL with state-of-the-art MLL, PLL, and PML approaches on five evaluation metrics, where the best performances are shown in bold face. (r = 1).

	Hamming Loss (the lower the better)							
Data set	Emotions	Genbase	Medical	Image	Corel5k	Bibtext	Eurlex-dc	Eurlex-sm
NATAL	**0.191 ± 0.026**	**0.002 ± 0.001**	**0.010 ± 0.002**	**0.172 ± 0.023**	0.011 ± 0.000	**0.009 ± 0.000**	0.005 ± 0.000	0.008 ± 0.001
ML-KNN	0.203 ± 0.017	0.005 ± 0.003	0.021 ± 0.001	0.182 ± 0.013	**0.009 ± 0.000**	0.015 ± 0.000	**0.002 ± 0.000**	0.008 ± 0.001
RankSVM	0.276 ± 0.054	0.014 ± 0.005	0.021 ± 0.002	0.309 ± 0.033	0.010 ± 0.002	0.021 ± 0.001	0.003 ± 0.001	0.009 ± 0.005
IPAL	0.274 ± 0.016	0.018 ± 0.003	0.017 ± 0.002	0.189 ± 0.012	0.010 ± 0.000	0.015 ± 0.000	**0.002 ± 0.000**	0.007 ± 0.002
LALO	0.311 ± 0.017	0.015 ± 0.004	0.025 ± 0.002	0.222 ± 0.007	0.019 ± 0.003	0.026 ± 0.002	0.013 ± 0.005	0.016 ± 0.005
PML-fp	0.300 ± 0.035	0.026 ± 0.010	0.047 ± 0.002	0.381 ± 0.013	0.012 ± 0.001	0.018 ± 0.000	0.010 ± 0.003	0.013 ± 0.002
PARTICLE	0.224 ± 0.024	0.012 ± 0.006	0.021 ± 0.002	0.206 ± 0.047	0.010 ± 0.001	0.016 ± 0.001	0.003 ± 0.000	**0.006 ± 0.000**
fPML	0.394 ± 0.021	0.005 ± 0.002	0.013 ± 0.003	0.250 ± 0.020	**0.009 ± 0.000**	0.013 ± 0.000	0.006 ± 0.002	0.010 ± 0.003
DRAMA	0.258 ± 0.020	0.003 ± 0.005	0.017 ± 0.003	0.194 ± 0.023	0.013 ± 0.002	0.010 ± 0.000	0.004 ± 0.001	0.008 ± 0.001
	Ranking Loss (the lower the better)							
Data set	Emotions	Genbase	Medical	Image	Corel5k	Bibtext	Eurlex-dc	Eurlex-sm
NATAL	**0.161 ± 0.028**	0.006 ± 0.014	**0.023 ± 0.013**	0.215 ± 0.016	0.277 ± 0.010	**0.069 ± 0.006**	**0.055 ± 0.005**	0.050 ± 0.005
ML-KNN	0.170 ± 0.022	0.006 ± 0.006	0.055 ± 0.011	0.205 ± 0.230	0.133 ± 0.006	0.243 ± 0.007	0.064 ± 0.003	0.050 ± 0.005
RankSVM	0.244 ± 0.092	0.005 ± 0.005	0.055 ± 0.012	0.247 ± 0.052	**0.112 ± 0.006**	0.224 ± 0.001	0.120 ± 0.005	0.079 ± 0.006
IPAL	0.657 ± 0.034	0.094 ± 0.022	0.361 ± 0.036	0.419 ± 0.023	0.912 ± 0.007	0.692 ± 0.006	0.305 ± 0.010	0.512 ± 0.007
LALO	0.198 ± 0.026	**0.004 ± 0.005**	0.071 ± 0.033	**0.147 ± 0.010**	0.268 ± 0.010	0.705 ± 0.003	0.625 ± 0.015	0.613 ± 0.008
PML-fp	0.285 ± 0.060	0.008 ± 0.003	0.043 ± 0.012	0.382 ± 0.024	0.152 ± 0.016	0.336 ± 0.011	0.075 ± 0.018	0.076 ± 0.009
PARTICLE	0.241 ± 0.022	0.022 ± 0.015	0.123 ± 0.032	0.218 ± 0.060	0.354 ± 0.055	0.307 ± 0.009	0.058 ± 0.004	**0.049 ± 0.002**
fPML	0.423 ± 0.043	0.009 ± 0.008	0.044 ± 0.016	0.224 ± 0.027	0.138 ± 0.006	0.092 ± 0.007	0.072 ± 0.002	0.065 ± 0.008
DRAMA	0.264 ± 0.031	0.006 ± 0.009	0.057 ± 0.016	0.188 ± 0.025	0.185 ± 0.003	0.192 ± 0.012	0.062 ± 0.009	0.062 ± 0.004
	One Error (the lower the better)							
Data set	Emotions	Genbase	Medical	Image	Corel5k	Bibtext	Eurlex-dc	Eurlex-sm
NATAL	**0.273 ± 0.087**	0.007 ± 0.009	**0.177 ± 0.045**	0.380 ± 0.030	**0.632 ± 0.026**	**0.375 ± 0.024**	0.352 ± 0.009	0.230 ± 0.016
ML-KNN	0.327 ± 0.060	0.021 ± 0.026	0.366 ± 0.042	0.342 ± 0.026	0.715 ± 0.017	0.723 ± 0.009	0.413 ± 0.010	0.230 ± 0.016
RankSVM	0.387 ± 0.134	0.056 ± 0.023	0.375 ± 0.031	0.448 ± 0.063	0.758 ± 0.013	0.518 ± 0.003	0.581 ± 0.021	0.241 ± 0.009
IPAL	0.385 ± 0.034	0.005 ± 0.007	0.260 ± 0.034	0.333 ± 0.021	0.696 ± 0.023	0.435 ± 0.003	**0.200 ± 0.007**	**0.136 ± 0.006**
LALO	0.267 ± 0.161	0.030 ± 0.034	0.236 ± 0.060	**0.040 ± 0.084**	0.661 ± 0.021	0.402 ± 0.005	0.235 ± 0.012	0.159 ± 0.002
PML-fp	0.349 ± 0.046	0.005 ± 0.003	0.265 ± 0.028	0.533 ± 0.034	0.732 ± 0.025	0.465 ± 0.010	0.412 ± 0.009	0.283 ± 0.015
PARTICLE	0.290 ± 0.049	0.015 ± 0.012	0.365 ± 0.048	0.336 ± 0.102	0.812 ± 0.075	0.575 ± 0.013	0.376 ± 0.009	0.230 ± 0.027
fPML	0.561 ± 0.052	**0.003 ± 0.006**	0.186 ± 0.041	0.444 ± 0.058	0.649 ± 0.024	0.406 ± 0.015	0.432 ± 0.014	0.252 ± 0.016
DRAMA	0.383 ± 0.059	0.009 ± 0.015	0.253 ± 0.038	0.342 ± 0.031	0.679 ± 0.026	0.402 ± 0.012	0.392 ± 0.012	0.243 ± 0.012
	Coverage (the lower the better)							
Data set	Emotions	Genbase	Medical	Image	Corel5k	Bibtext	Eurlex-dc	Eurlex-sm
NATAL	**0.288 ± 0.025**	**0.016 ± 0.023**	**0.035 ± 0.016**	**0.189 ± 0.013**	0.576 ± 0.013	**0.128 ± 0.008**	**0.063 ± 0.006**	**0.087 ± 0.006**
ML-KNN	0.304 ± 0.026	0.021 ± 0.011	0.075 ± 0.016	0.218 ± 0.022	0.335 ± 0.011	0.382 ± 0.012	0.081 ± 0.004	0.088 ± 0.006
RankSVM	0.372 ± 0.079	0.026 ± 0.005	0.074 ± 0.015	0.253 ± 0.044	0.435 ± 0.012	0.276 ± 0.013	0.149 ± 0.031	0.356 ± 0.012
IPAL	0.483 ± 0.032	0.101 ± 0.029	0.203 ± 0.025	0.258 ± 0.016	0.909 ± 0.005	0.511 ± 0.013	0.271 ± 0.011	0.591 ± 0.007
LALO	0.749 ± 0.069	0.036 ± 0.012	0.076 ± 0.022	0.875 ± 0.043	0.698 ± 0.011	0.498 ± 0.012	0.299 ± 0.010	0.698 ± 0.015
PML-fp	0.425 ± 0.056	0.025 ± 0.008	0.046 ± 0.019	0.361 ± 0.023	0.532 ± 0.021	0.325 ± 0.013	0.109 ± 0.013	0.153 ± 0.006
PARTICLE	0.362 ± 0.040	0.042 ± 0.025	0.125 ± 0.028	0.213 ± 0.067	0.558 ± 0.059	0.469 ± 0.012	0.094 ± 0.004	0.110 ± 0.004
fPML	0.511 ± 0.028	0.030 ± 0.017	0.064 ± 0.025	0.231 ± 0.024	**0.321 ± 0.011**	0.163 ± 0.011	0.108 ± 0.012	0.108 ± 0.006
DRAMA	0.381 ± 0.043	0.025 ± 0.016	0.072 ± 0.021	0.196 ± 0.029	0.465 ± 0.015	0.198 ± 0.015	0.075 ± 0.016	0.092 ± 0.005
	Average Precision (the higher the better)							
Data set	Emotions	Genbase	Medical	Image	Corel5k	Bibtext	Eurlex-dc	Eurlex-sm
NATAL	**0.812 ± 0.036**	**0.995 ± 0.022**	**0.870 ± 0.035**	0.750 ± 0.016	**0.278 ± 0.014**	**0.591 ± 0.016**	**0.711 ± 0.003**	0.764 ± 0.013
ML-KNN	0.793 ± 0.020	0.980 ± 0.018	0.717 ± 0.021	0.769 ± 0.020	0.255 ± 0.007	0.260 ± 0.007	0.635 ± 0.008	**0.794 ± 0.005**
RankSVM	0.724 ± 0.087	0.965 ± 0.014	0.720 ± 0.022	0.709 ± 0.046	0.265 ± 0.008	0.325 ± 0.008	0.449 ± 0.015	0.532 ± 0.012
IPAL	0.648 ± 0.023	0.920 ± 0.019	0.687 ± 0.032	0.750 ± 0.014	0.099 ± 0.007	0.343 ± 0.004	0.697 ± 0.010	0.497 ± 0.007
LALO	0.667 ± 0.031	0.648 ± 0.062	0.412 ± 0.019	0.740 ± 0.018	0.193 ± 0.009	0.312 ± 0.008	0.625 ± 0.011	0.436 ± 0.006
PML-fp	0.710 ± 0.064	0.981 ± 0.033	0.721 ± 0.031	0.617 ± 0.022	0.260 ± 0.009	0.325 ± 0.011	0.637 ± 0.012	0.609 ± 0.016
PARTICLE	0.758 ± 0.023	0.972 ± 0.020	0.661 ± 0.041	0.764 ± 0.063	0.167 ± 0.046	0.291 ± 0.010	0.674 ± 0.083	0.678 ± 0.014
fPML	0.577 ± 0.032	0.985 ± 0.012	0.845 ± 0.035	0.719 ± 0.032	**0.276 ± 0.009**	0.542 ± 0.012	0.663 ± 0.022	0.735 ± 0.012
DRAMA	0.705 ± 0.028	0.986 ± 0.027	0.801 ± 0.035	**0.781 ± 0.019**	0.234 ± 0.006	0.534 ± 0.012	0.682 ± 0.021	0.749 ± 0.012

4.3. Ablation study and robustness analysis

In order to demonstrate the effectiveness of our employed *Feature Completion* (FC) operation, we conduct the **Ablation Study** on our proposed NATAL. Specifically, we compare the NATAL algorithm (FC operation) with its degenerated algorithm **OP (2)** (FC-free operation) that does not contain FC operation on all employed PML data sets. The detailed experimental results are summarized in Table 7. According to Table 7, it is observed that our proposed NATAL algorithm significantly outperforms FC-free algorithm in all cases, which demonstrates the validity of the FC operation.

Table 4
Comparison of NATAL with state-of-the-art MLL, PLL, and PML approaches on five evaluation metrics, where the best performances are shown in bold face. (r = 2).

	Hamming Loss (the lower the better)							
Data set	Emotions	Genbase	Medical	Image	Corel5k	Bibtext	Eurlex-dc	Eurlex-sm
NATAL	**0.228 ± 0.024**	0.003 ± 0.002	**0.012 ± 0.003**	0.220 ± 0.012	**0.008 ± 0.000**	**0.010 ± 0.006**	0.012 ± 0.003	**0.006 ± 0.010**
ML-KNN	0.258 ± 0.011	0.005 ± 0.002	0.021 ± 0.001	0.683 ± 0.026	0.009 ± 0.000	0.013 ± 0.001	0.010 ± 0.002	0.008 ± 0.002
RankSVM	0.363 ± 0.074	0.021 ± 0.005	0.053 ± 0.002	0.559 ± 0.017	0.012 ± 0.001	0.025 ± 0.002	0.005 ± 0.002	0.011 ± 0.004
IPAL	0.286 ± 0.019	0.018 ± 0.003	0.017 ± 0.001	0.189 ± 0.013	0.010 ± 0.000	0.014 ± 0.001	**0.003 ± 0.001**	0.007 ± 0.000
LALO	0.311 ± 0.010	0.015 ± 0.003	0.027 ± 0.001	0.235 ± 0.008	0.021 ± 0.002	0.023 ± 0.001	0.009 ± 0.003	0.018 ± 0.004
PML-fp	0.392 ± 0.023	0.036 ± 0.002	0.055 ± 0.005	0.424 ± 0.023	0.012 ± 0.001	0.017 ± 0.002	0.007 ± 0.005	0.015 ± 0.004
PARTICLE	0.229 ± 0.017	0.010 ± 0.005	0.020 ± 0.003	0.240 ± 0.043	0.009 ± 0.000	0.016 ± 0.001	**0.003 ± 0.001**	0.009 ± 0.001
fPML	0.430 ± 0.037	0.003 ± 0.001	**0.012 ± 0.006**	0.257 ± 0.019	0.009 ± 0.000	0.013 ± 0.000	0.008 ± 0.005	0.012 ± 0.004
DRAMA	0.289 ± 0.028	**0.002 ± 0.004**	0.016 ± 0.002	**0.187 ± 0.012**	0.015 ± 0.003	0.012 ± 0.001	0.006 ± 0.002	0.010 ± 0.002
	Ranking Loss (the lower the better)							
Data set	Emotions	Genbase	Medical	Image	Corel5k	Bibtext	Eurlex-dc	Eurlex-sm
NATAL	**0.190 ± 0.027**	**0.001 ± 0.001**	**0.025 ± 0.007**	0.256 ± 0.014	0.278 ± 0.001	**0.073 ± 0.003**	**0.059 ± 0.002**	0.041 ± 0.005
ML-KNN	0.204 ± 0.018	0.008 ± 0.003	0.072 ± 0.009	0.241 ± 0.023	**0.135 ± 0.007**	0.223 ± 0.007	0.072 ± 0.005	**0.032 ± 0.001**
RankSVM	0.225 ± 0.065	0.006 ± 0.004	0.086 ± 0.001	0.230 ± 0.036	0.165 ± 0.008	0.243 ± 0.008	0.132 ± 0.006	0.082 ± 0.003
IPAL	0.680 ± 0.042	0.214 ± 0.017	0.375 ± 0.026	0.440 ± 0.028	0.914 ± 0.008	0.695 ± 0.034	0.316 ± 0.008	0.510 ± 0.005
LALO	0.221 ± 0.017	0.002 ± 0.004	0.075 ± 0.036	0.166 ± 0.019	0.271 ± 0.009	0.710 ± 0.005	0.635 ± 0.012	0.621 ± 0.009
PML-fp	0.401 ± 0.035	0.009 ± 0.001	0.050 ± 0.010	0.426 ± 0.043	0.162 ± 0.013	0.341 ± 0.009	0.079 ± 0.016	0.082 ± 0.012
PARTICLE	0.252 ± 0.028	0.025 ± 0.016	0.099 ± 0.025	0.250 ± 0.048	0.349 ± 0.070	0.287 ± 0.011	0.062 ± 0.004	0.053 ± 0.001
fPML	0.377 ± 0.078	0.007 ± 0.005	0.043 ± 0.011	0.238 ± 0.021	0.137 ± 0.004	0.087 ± 0.004	0.078 ± 0.001	0.068 ± 0.008
DRAMA	0.265 ± 0.031	0.008 ± 0.003	0.049 ± 0.026	**0.181 ± 0.024**	0.192 ± 0.012	0.203 ± 0.011	0.068 ± 0.008	0.051 ± 0.006
	One Error (the lower the better)							
Data set	Emotions	Genbase	Medical	Image	Corel5k	Bibtext	Eurlex-dc	Eurlex-sm
NATAL	0.301 ± 0.039	**0.000 ± 0.000**	**0.182 ± 0.033**	0.426 ± 0.023	**0.654 ± 0.027**	**0.373 ± 0.020**	0.368 ± 0.009	0.215 ± 0.013
ML-KNN	0.319 ± 0.075	0.024 ± 0.022	0.383 ± 0.035	0.406 ± 0.040	0.724 ± 0.021	0.620 ± 0.024	0.469 ± 0.013	0.186 ± 0.004
RankSVM	0.371 ± 0.097	0.053 ± 0.038	0.532 ± 0.043	0.416 ± 0.055	0.768 ± 0.012	0.529 ± 0.016	0.589 ± 0.012	0.246 ± 0.010
IPAL	0.424 ± 0.043	0.126 ± 0.045	0.276 ± 0.028	0.355 ± 0.027	0.704 ± 0.027	0.436 ± 0.062	**0.209 ± 0.005**	**0.127 ± 0.008**
LALO	**0.267 ± 0.013**	0.036 ± 0.044	0.261 ± 0.010	**0.040 ± 0.084**	0.663 ± 0.019	0.410 ± 0.019	0.226 ± 0.011	0.162 ± 0.009
PML-fp	0.476 ± 0.067	**0.000 ± 0.000**	0.282 ± 0.053	0.619 ± 0.054	0.746 ± 0.021	0.435 ± 0.012	0.405 ± 0.003	0.286 ± 0.016
PARTICLE	0.293 ± 0.065	0.003 ± 0.006	0.245 ± 0.045	0.389 ± 0.088	0.823 ± 0.091	0.549 ± 0.015	0.357 ± 0.007	0.244 ± 0.006
fPML	0.558 ± 0.061	0.002 ± 0.005	0.196 ± 0.036	0.457 ± 0.026	0.672 ± 0.030	0.406 ± 0.021	0.441 ± 0.012	0.249 ± 0.013
DRAMA	0.383 ± 0.089	0.001 ± 0.015	0.249 ± 0.012	0.334 ± 0.016	0.680 ± 0.012	0.413 ± 0.012	0.401 ± 0.008	0.236 ± 0.009
	Coverage (the lower the better)							
Data set	Emotions	Genbase	Medical	Image	Corel5k	Bibtext	Eurlex-dc	Eurlex-sm
NATAL	**0.331 ± 0.033**	**0.011 ± 0.004**	**0.037 ± 0.009**	0.257 ± 0.015	0.579 ± 0.016	**0.135 ± 0.005**	**0.068 ± 0.013**	**0.085 ± 0.003**
ML-KNN	0.346 ± 0.030	0.024 ± 0.009	0.097 ± 0.014	0.247 ± 0.018	**0.310 ± 0.014**	0.358 ± 0.013	0.102 ± 0.005	0.086 ± 0.006
RankSVM	0.352 ± 0.053	0.017 ± 0.007	0.105 ± 0.016	0.239 ± 0.029	0.445 ± 0.035	0.285 ± 0.011	0.152 ± 0.008	0.359 ± 0.011
IPAL	0.494 ± 0.038	0.017 ± 0.007	0.105 ± 0.016	0.239 ± 0.029	0.445 ± 0.035	0.285 ± 0.011	0.152 ± 0.008	0.359 ± 0.011
LALO	0.793 ± 0.056	0.033 ± 0.004	0.082 ± 0.022	0.892 ± 0.054	0.695 ± 0.008	0.502 ± 0.003	0.236 ± 0.011	0.703 ± 0.016
PML-fp	0.487 ± 0.052	0.028 ± 0.008	0.050 ± 0.033	0.395 ± 0.033	0.436 ± 0.016	0.333 ± 0.015	0.091 ± 0.012	0.155 ± 0.009
PARTICLE	0.366 ± 0.046	0.046 ± 0.030	0.115 ± 0.028	0.234 ± 0.065	0.561 ± 0.063	0.450 ± 0.014	0.097 ± 0.007	0.113 ± 0.003
fPML	0.471 ± 0.054	**0.011 ± 0.011**	0.063 ± 0.016	0.247 ± 0.020	**0.310 ± 0.008**	0.156 ± 0.006	0.099 ± 0.008	0.100 ± 0.007
DRAMA	0.378 ± 0.046	0.026 ± 0.015	0.063 ± 0.012	**0.197 ± 0.023**	0.468 ± 0.012	0.205 ± 0.011	0.081 ± 0.012	0.096 ± 0.008
	Average Precision (the higher the better)							
Data set	Emotions	Genbase	Medical	Image	Corel5k	Bibtext	Eurlex-dc	Eurlex-sm
NATAL	**0.768 ± 0.026**	**0.997 ± 0.004**	**0.858 ± 0.024**	0.714 ± 0.010	**0.272 ± 0.012**	**0.588 ± 0.013**	**0.700 ± 0.017**	0.768 ± 0.006
ML-KNN	0.765 ± 0.023	0.977 ± 0.010	0.701 ± 0.021	0.729 ± 0.024	0.252 ± 0.011	0.319 ± 0.015	0.625 ± 0.009	**0.773 ± 0.005**
RankSVM	0.735 ± 0.062	0.964 ± 0.025	0.599 ± 0.024	0.729 ± 0.036	0.250 ± 0.045	0.356 ± 0.005	0.432 ± 0.009	0.528 ± 0.011
IPAL	0.630 ± 0.030	0.807 ± 0.045	0.673 ± 0.029	0.734 ± 0.018	0.097 ± 0.008	0.340 ± 0.035	0.672 ± 0.005	0.499 ± 0.006
LALO	0.639 ± 0.028	0.653 ± 0.072	0.404 ± 0.030	0.708 ± 0.026	0.183 ± 0.007	0.339 ± 0.012	0.611 ± 0.009	0.442 ± 0.008
PML-fp	0.618 ± 0.240	0.986 ± 0.004	0.706 ± 0.016	0.563 ± 0.038	0.250 ± 0.009	0.319 ± 0.009	0.618 ± 0.006	0.600 ± 0.016
PARTICLE	0.749 ± 0.029	0.978 ± 0.016	0.756 ± 0.041	0.728 ± 0.054	0.162 ± 0.054	0.315 ± 0.012	0.631 ± 0.009	0.667 ± 0.028
fPML	0.605 ± 0.049	0.989 ± 0.006	0.852 ± 0.031	0.708 ± 0.017	0.268 ± 0.013	0.544 ± 0.011	0.658 ± 0.019	0.729 ± 0.013
DRAMA	0.716 ± 0.046	0.986 ± 0.019	0.811 ± 0.026	**0.783 ± 0.021**	0.228 ± 0.004	0.549 ± 0.008	0.675 ± 0.010	0.752 ± 0.013

In addition, to demonstrate the robustness of our proposed NATAL, we conduct another group of comparative experiments, where the proportion of training examples decreases from 90% to 10%. Note that, it is the first time to evaluate the robustness of PML algorithm w.r.t the number of training examples. Fig. 4 illustrates the comparative results between NATAL and other methods on five evaluation metrics. As described in Fig. 4, NATAL is superior to all other comparing methods on most evaluation metrics, and only partially inferior to IPAL (that is based on PLL framework) on *One Error* metric. Especially, when the scale of training examples is small (10%), NATAL can significantly outperform all comparing methods on each evaluation metric. In summary, the robustness of our proposed method is demonstrated.

Table 5
Comparison of NATAL with state-of-the-art MLL, PLL, and PML approaches on five evaluation metrics, where the best performances are shown in bold face. (r = 3).

	Hamming Loss (the lower the better)							
Data set	Emotions	Genbase	Medical	Image	Corel5k	Bibtext	Eurlex-dc	Eurlex-sm
NATAL	**0.264 ± 0.017**	**0.005 ± 0.002**	**0.015 ± 0.006**	0.308 ± 0.017	0.011 ± 0.000	**0.012 ± 0.008**	0.015 ± 0.002	0.012 ± 0.001
ML-KNN	0.607 ± 0.029	0.006 ± 0.002	0.022 ± 0.002	0.753 ± 0.005	0.009 ± 0.000	0.014 ± 0.000	0.009 ± 0.005	0.012 ± 0.001
RankSVM	0.517 ± 0.020	0.030 ± 0.008	0.074 ± 0.003	0.650 ± 0.020	0.019 ± 0.001	0.021 ± 0.000	0.006 ± 0.001	0.009 ± 0.006
IPAL	0.314 ± 0.021	0.021 ± 0.008	0.018 ± 0.003	0.201 ± 0.008	0.010 ± 0.000	0.015 ± 0.000	**0.002 ± 0.000**	0.009 ± 0.002
LALO	0.311 ± 0.002	0.016 ± 0.004	0.027 ± 0.001	0.244 ± 0.005	0.023 ± 0.000	0.020 ± 0.000	0.010 ± 0.004	0.013 ± 0.005
PML-fp	0.437 ± 0.027	0.043 ± 0.012	0.056 ± 0.005	0.448 ± 0.030	0.010 ± 0.001	0.019 ± 0.000	0.011 ± 0.006	0.016 ± 0.002
PARTICLE	0.233 ± 0.018	0.010 ± 0.005	0.024 ± 0.002	0.403 ± 0.042	**0.009 ± 0.000**	0.017 ± 0.000	0.004 ± 0.000	0.010 ± 0.001
fPML	0.394 ± 0.030	**0.005 ± 0.002**	0.016 ± 0.002	0.257 ± 0.005	**0.009 ± 0.000**	0.013 ± 0.000	0.007 ± 0.005	**0.005 ± 0.002**
DRAMA	0.339 ± 0.019	0.005 ± 0.003	0.018 ± 0.001	**0.203 ± 0.006**	0.013 ± 0.002	0.016 ± 0.000	0.005 ± 0.001	0.008 ± 0.001
	Ranking Loss (the lower the better)							
Data set	Emotions	Genbase	Medical	Image	Corel5k	Bibtext	Eurlex-dc	Eurlex-sm
NATAL	**0.227 ± 0.030**	**0.008 ± 0.031**	**0.025 ± 0.006**	0.343 ± 0.028	0.285 ± 0.012	**0.078 ± 0.005**	**0.049 ± 0.019**	0.045 ± 0.012
ML-KNN	0.241 ± 0.026	0.009 ± 0.005	0.088 ± 0.019	0.342 ± 0.026	0.139 ± 0.007	0.232 ± 0.006	0.086 ± 0.012	**0.043 ± 0.003**
RankSVM	0.235 ± 0.037	**0.008 ± 0.005**	0.103 ± 0.018	0.245 ± 0.023	0.170 ± 0.012	0.224 ± 0.006	0.134 ± 0.012	0.085 ± 0.005
IPAL	0.738 ± 0.052	0.246 ± 0.107	0.383 ± 0.055	0.469 ± 0.018	0.917 ± 0.008	0.689 ± 0.018	0.326 ± 0.005	0.506 ± 0.006
LALO	0.240 ± 0.024	0.009 ± 0.018	0.079 ± 0.035	**0.196 ± 0.025**	0.275 ± 0.008	0.703 ± 0.009	0.663 ± 0.011	0.619 ± 0.011
PML-fp	0.462 ± 0.034	0.035 ± 0.016	0.052 ± 0.016	0.494 ± 0.043	0.165 ± 0.012	0.336 ± 0.002	0.068 ± 0.014	0.079 ± 0.012
PARTICLE	0.259 ± 0.019	0.026 ± 0.017	0.100 ± 0.021	0.315 ± 0.073	0.333 ± 0.050	0.287 ± 0.010	0.061 ± 0.023	0.053 ± 0.002
fPML	0.434 ± 0.033	0.012 ± 0.006	0.040 ± 0.019	0.238 ± 0.013	**0.136 ± 0.005**	0.092 ± 0.007	0.075 ± 0.003	0.071 ± 0.007
DRAMA	0.298 ± 0.026	0.010 ± 0.025	0.032 ± 0.019	0.221 ± 0.023	0.195 ± 0.011	0.189 ± 0.009	0.062 ± 0.010	0.055 ± 0.007
	One Error (the lower the better)							
Data set	Emotions	Genbase	Medical	Image	Corel5k	Bibtext	Eurlex-dc	Eurlex-sm
NATAL	0.390 ± 0.035	0.016 ± 0.030	**0.173 ± 0.023**	0.553 ± 0.044	**0.659 ± 0.026**	**0.378 ± 0.016**	0.345 ± 0.017	0.227 ± 0.008
ML-KNN	0.383 ± 0.062	0.021 ± 0.016	0.425 ± 0.041	0.562 ± 0.035	0.718 ± 0.020	0.624 ± 0.008	0.502 ± 0.015	0.189 ± 0.003
RankSVM	0.389 ± 0.096	0.064 ± 0.048	0.588 ± 0.056	0.437 ± 0.030	0.770 ± 0.023	0.518 ± 0.012	0.602 ± 0.021	0.539 ± 0.012
IPAL	0.511 ± 0.093	0.158 ± 0.118	0.285 ± 0.059	0.386 ± 0.017	0.712 ± 0.028	0.443 ± 0.018	0.213 ± 0.002	**0.129 ± 0.009**
LALO	**0.300 ± 0.153**	0.047 ± 0.047	0.285 ± 0.096	**0.120 ± 0.140**	0.668 ± 0.023	0.405 ± 0.020	**0.244 ± 0.015**	0.172 ± 0.012
PML-fp	0.527 ± 0.049	0.012 ± 0.005	0.295 ± 0.035	0.712 ± 0.053	0.747 ± 0.035	0.465 ± 0.013	0.429 ± 0.008	0.292 ± 0.015
PARTICLE	0.306 ± 0.065	**0.003 ± 0.006**	0.251 ± 0.066	0.410 ± 0.121	0.855 ± 0.060	0.557 ± 0.014	0.356 ± 0.012	0.212 ± 0.017
fPML	0.547 ± 0.043	**0.003 ± 0.010**	0.176 ± 0.028	0.443 ± 0.029	0.664 ± 0.013	0.406 ± 0.015	0.439 ± 0.010	0.265 ± 0.011
DRAMA	0.483 ± 0.067	0.015 ± 0.050	0.242 ± 0.049	0.359 ± 0.028	0.675 ± 0.015	0.398 ± 0.012	0.382 ± 0.012	0.218 ± 0.009
	Coverage (the lower the better)							
Data set	Emotions	Genbase	Medical	Image	Corel5k	Bibtext	Eurlex-dc	Eurlex-sm
NATAL	**0.356 ± 0.031**	0.027 ± 0.031	**0.036 ± 0.011**	0.323 ± 0.024	0.581 ± 0.017	**0.143 ± 0.008**	**0.060 ± 0.021**	**0.089 ± 0.005**
ML-KNN	0.374 ± 0.028	0.025 ± 0.009	0.111 ± 0.027	0.324 ± 0.020	0.321 ± 0.012	0.369 ± 0.009	0.098 ± 0.002	**0.089 ± 0.008**
RankSVM	0.368 ± 0.040	0.022 ± 0.008	0.125 ± 0.021	0.253 ± 0.021	0.435 ± 0.014	0.294 ± 0.008	0.158 ± 0.021	0.339 ± 0.011
IPAL	0.514 ± 0.036	0.231 ± 0.090	0.220 ± 0.031	0.283 ± 0.017	0.912 ± 0.005	0.499 ± 0.004	0.291 ± 0.013	0.602 ± 0.004
LALO	0.815 ± 0.064	0.037 ± 0.011	0.084 ± 0.026	0.917 ± 0.042	0.699 ± 0.005	0.498 ± 0.023	0.239 ± 0.012	0.689 ± 0.012
PML-fp	0.528 ± 0.035	0.039 ± 0.012	0.055 ± 0.021	0.448 ± 0.032	0.450 ± 0.024	0.356 ± 0.016	0.121 ± 0.019	0.166 ± 0.011
PARTICLE	0.365 ± 0.040	0.049 ± 0.031	0.121 ± 0.023	0.275 ± 0.088	0.550 ± 0.066	0.458 ± 0.014	0.101 ± 0.021	0.114 ± 0.004
fPML	0.534 ± 0.028	0.032 ± 0.011	0.059 ± 0.026	0.250 ± 0.015	**0.317 ± 0.010**	0.163 ± 0.011	0.095 ± 0.009	0.103 ± 0.015
DRAMA	0.422 ± 0.038	**0.011 ± 0.016**	0.040 ± 0.028	**0.230 ± 0.019**	0.465 ± 0.021	0.235 ± 0.010	0.078 ± 0.013	0.092 ± 0.005
	Average Precision (the higher the better)							
Data set	Emotions	Genbase	Medical	Image	Corel5k	Bibtext	Eurlex-dc	Eurlex-sm
NATAL	0.728 ± 0.024	**0.985 ± 0.031**	**0.854 ± 0.020**	0.633 ± 0.029	0.265 ± 0.015	**0.580 ± 0.012**	**0.716 ± 0.016**	**0.762 ± 0.011**
ML-KNN	0.741 ± 0.029	0.975 ± 0.014	0.672 ± 0.039	0.627 ± 0.022	0.251 ± 0.007	0.306 ± 0.006	0.603 ± 0.012	0.761 ± 0.012
RankSVM	0.722 ± 0.050	0.955 ± 0.027	0.544 ± 0.044	0.714 ± 0.021	0.243 ± 0.007	0.329 ± 0.002	0.413 ± 0.015	0.530 ± 0.012
IPAL	0.588 ± 0.051	0.774 ± 0.105	0.667 ± 0.048	0.717 ± 0.013	0.093 ± 0.008	0.347 ± 0.016	0.670 ± 0.009	0.490 ± 0.005
LALO	0.608 ± 0.045	0.641 ± 0.084	0.404 ± 0.036	0.658 ± 0.025	0.179 ± 0.002	0.326 ± 0.012	0.603 ± 0.009	0.436 ± 0.009
PML-fp	0.573 ± 0.025	0.973 ± 0.023	0.700 ± 0.013	0.501 ± 0.037	0.221 ± 0.010	0.297 ± 0.008	0.613 ± 0.009	0.605 ± 0.011
PARTICLE	**0.745 ± 0.024**	0.976 ± 0.017	0.720 ± 0.044	0.689 ± 0.096	0.135 ± 0.041	0.313 ± 0.012	0.630 ± 0.016	0.695 ± 0.012
fPML	0.573 ± 0.024	0.983 ± 0.009	0.853 ± 0.027	0.712 ± 0.014	**0.271 ± 0.007**	0.542 ± 0.012	0.659 ± 0.015	0.731 ± 0.011
DRAMA	0.722 ± 0.050	0.981 ± 0.008	0.828 ± 0.032	**0.759 ± 0.031**	0.219 ± 0.003	0.536 ± 0.006	0.679 ± 0.011	0.749 ± 0.012

Table 6
Friedman statics τ_F in terms of each evaluation metric (at 0.05 significance level).

Evaluation Metric	Hamming Loss	Ranking Loss	One error	Coverage	Average Precision
τ_F	8.05	4.72	4.93	9.29	5.42
critical value	2.109 (Methods: 9, Data sets: 8)				

Table 7
The experimental comparisons between the proposed NATAL and FC-free algorithm on all PML data sets.

Evaluation Metric	Methods	Emotions	Genbase	Medical	Image	Corel5k	Bibtext	Eurlex-dc	Eurlex-sm
Hamming Loss	NATAL	**0.191 ± 0.026**	**0.002 ± 0.001**	**0.010 ± 0.002**	**0.172 ± 0.023**	**0.011 ± 0.000**	**0.009 ± 0.000**	**0.012 ± 0.000**	**0.018 ± 0.001**
	FC-free	0.296 ± 0.021	0.160 ± 0.002	0.161 ± 0.029	0.308 ± 0.003	0.011 ± 0.020	0.297 ± 0.017	0.322 ± 0.001	0.281 ± 0.001
Ranking Loss	NATAL	**0.161 ± 0.028**	**0.015 ± 0.002**	**0.023 ± 0.013**	**0.215 ± 0.016**	**0.277 ± 0.010**	**0.069 ± 0.006**	**0.055 ± 0.005**	**0.050 ± 0.005**
	FC-free	0.254 ± 0.035	0.120 ± 0.002	0.141 ± 0.028	0.293 ± 0.011	0.318 ± 0.009	0.309 ± 0.013	0.372 ± 0.007	0.374 ± 0.009
One Error	NATAL	**0.273 ± 0.087**	**0.015 ± 0.009**	**0.177 ± 0.045**	**0.380 ± 0.030**	**0.632 ± 0.026**	**0.375 ± 0.024**	**0.352 ± 0.009**	**0.230 ± 0.016**
	FC-free	0.361 ± 0.048	0.134 ± 0.012	0.242 ± 0.025	0.456 ± 0.013	0.662 ± 0.012	0.508 ± 0.016	0.589 ± 0.008	0.519 ± 0.015
Coverage	NATAL	**0.288 ± 0.025**	**0.024 ± 0.023**	**0.035 ± 0.016**	**0.189 ± 0.013**	**0.576 ± 0.013**	**0.128 ± 0.008**	**0.063 ± 0.006**	**0.087 ± 0.006**
	FC-free	0.669 ± 0.026	0.122 ± 0.021	0.152 ± 0.031	0.277 ± 0.003	0.636 ± 0.015	0.379 ± 0.016	0.376 ± 0.007	0.390 ± 0.010
Average Precision	NATAL	**0.812 ± 0.036**	**0.995 ± 0.022**	**0.870 ± 0.035**	**0.750 ± 0.016**	**0.278 ± 0.014**	**0.591 ± 0.016**	**0.711 ± 0.003**	**0.764 ± 0.013**
	FC-free	0.736 ± 0.020	0.878 ± 0.010	0.781 ± 0.025	0.691 ± 0.008	0.247 ± 0.013	0.450 ± 0.015	0.468 ± 0.008	0.476 ± 0.012

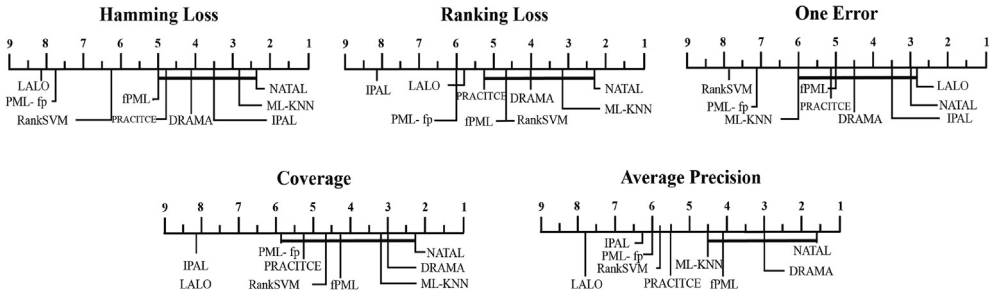

Fig. 3. Comparison of NATAL (control algorithm) against eight comparing algorithms with the Bonferroni-Dunn test. Algorithms not connected with NATAL in the CD diagram are considered to have significantly different performance from the control algorithm (CD = 3.73 at 0.05 significance level).

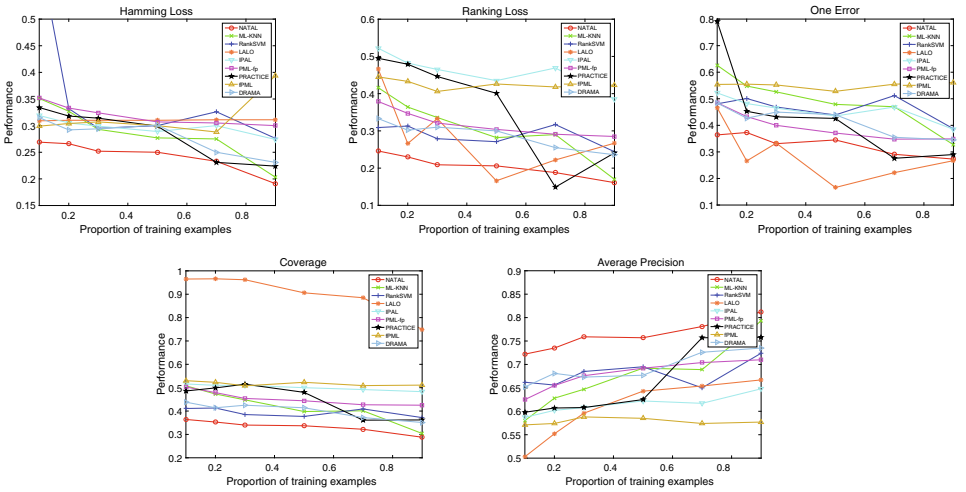

Fig. 4. The performance of each comparing method on *Emotions* data set changes as the proportion of training examples increases from 0.1 to 0.9 (with one false candidate label [$r = 1$]).

5. Further analysis

Complexity Analysis: At each iteration of Algorithm 1, the dominant computational cost comes from the calculation of the singular value decomposition (SVD) of $\mathbf{W}^t$ and $\mathbf{E}_X^t$. Given the feature matrix $\mathbf{X} \in \mathbb{R}^{d\times m}$ and label matrix $\mathbf{Y} \in \mathbb{R}^{q\times m}$, the cost complexity of SVD w.r.t $\mathbf{W}$ is $\mathcal{O}(qdr)$, where r denotes the rank of matrix $\mathbf{W}$ if we choose skinny SVD. Similarly, the cost complexity of SVD w.r.t $\mathbf{E}_X$ is $\mathcal{O}(mdr)$. Thus, the whole cost complexity of NATAL is $\mathcal{O}(T * (q + m)dr)$, where T is the number of iterations.

Parameter Analysis: We study the sensitivity analysis of NATAL with respect to its three employed parameters $\alpha,\ \beta,\ \lambda$. Fig. 5 shows the performance of NATAL under different parameter configurations on *Emotions* data set. According to Fig. 5, the parameters usually follow the optimal configurations ($\alpha = 1,\ \beta = 10^{-6},\ \lambda = 0.1$) but vary with minor adjustments on different data sets.

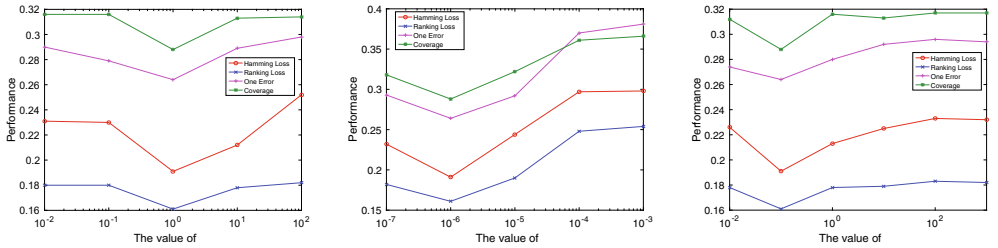

Fig. 5. Parameter Analysis: The performance of NATAL changes as each parameter increases with other parameters fixed.

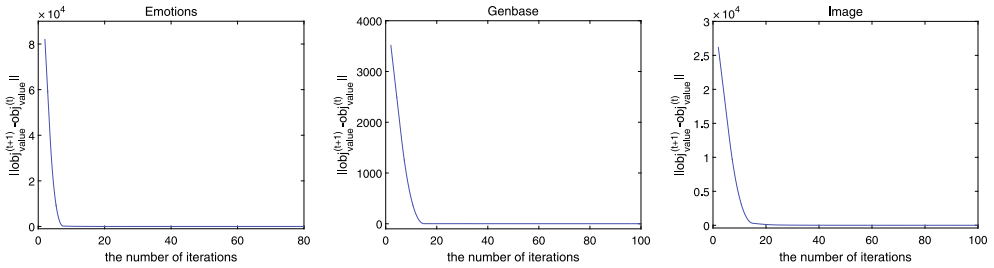

Fig. 6. The convergence curves of NATAL on *Emotions*, *Genbase* and *Image* data sets with increasing number of iterations.

Convergence Analysis: We conduct the convergence analysis of NATAL on *Emotions*, *Genbase* and *Image* data sets, where the convergence curves are separately shown in the sub-figures of Fig. 6. We can easily observe that each $\|obj_{value}^{(t+1)} - obj_{value}^{(t)}\|$ gradually decreases to 0 as the number of iterations t increases. Therefore, the convergence of NATAL is demonstrated.

6. Conclusion

In this paper, we propose a novel noisy label tolerated PML framework named NATAL, which provides a new perspective for solving PML problem. In NATAL, the label information is considered to be precise while the feature information is assumed to be missing. Therefore, we transfer such label redundancy problem to a feature missing problem, and solve it via *Feature Completion* (FC) scheme. Ablation Study proves the effectiveness of such FC operation and enormous experiments demonstrate that NATAL can achieve superior performance against state-of-the-art methods.

CRediT authorship contribution statement

Gengyu Lyu: Data curation, Software, Writing - original draft. **Songhe Feng:** Methodology, Supervision. **Yidong Li:** 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.

Acknowledgement

This work was supported in part by National Key Research and Development Project (No. 2018AAA0100300), in part by the National Natural Science Foundation of China (No. 61872032), in part by the Beijing Natural Science Foundation (Nos. 4202058, 4202057, 4202060, 9192008), in part by the Fundamental Research Funds for the Central universities (2019JBM020, 2020YJS026), and in part by the Key R&D Program of Zhejiang Province (Nos. 2019C01068).

References

[1] M. Xie, S. Huang, Partial multi-label learning, in: AAAI Conference on Artificial Intelligence, 2018, pp. 4302–4309..
[2] G. Yu, X. Chen, C. Domenicino, J. Wang, Z. Li, Z. Zhang, X. Wu, Featured-induced partial multi-label learning, in: International Conference on Data Mining, 2018, pp. 1398–1403.
[3] J. Fang, M. Zhang, Partial multi-label learning via credible label elicitation, in: AAAI Conference on Artificial Intelligence, 2019, pp. 3518–3525..
[4] M. Zhang, Z. Zhou, A review on multi-label learning algorithms, IEEE Transactions on Knowledge and Data Engineering 26 (8) (2013) 1819–1837.
[5] C. Chen, V. Patel, R. Chellappa, Learning from ambiguously labeled face images, IEEE Transactions on Pattern Analysis and Machine Intelligence 40 (7) (2017) 1653–1667.
[6] L. Feng, B. An, Partial label learning by semantic difference maxization., in: International Joint Conference on Artificial Intelligence, 2019, pp. 2294–2300..
[7] L. Sun, S. Feng, T. Wang, C. Lang, Y. Jin, Partial multi-label learning via low-rank and sparse decomposition, in: AAAI Conference on Artificial Intelligence, 2019, pp. 5016–5023..
[8] H. Wang, W. Liu, Y. Zhao, C. Zhang, T. Hu, G. Chen, Discriminative and correlative partial multi label learning, in: International Joint Conference on Artificial Intelligence, 2019, pp. 3691–3697..
[9] T. Cour, B. Sapp, B. Taskar, Learning from partial labels, IEEE Transactions on Knowledge and Data Engineering 12 (5) (2011) 1501–1536.
[10] M. Zhang, Y. Li, X. Liu, X. Geng, Binary relevance for multi-label learning: an overview, Frontiers of Computer Science 12 (2) (2018) 191–202.
[11] J. Fürnkranz, E. Hüllermeier, E. Mencía, K. Brinker, Multi label classification via calibrated label ranking, Machine Learning 73 (2) (2008) 133–153.
[12] N. Ueda, K. Saito, Parametric mixture models for multi-labeled text, in: Advances in Neural Information Processing Systems, 2003, pp. 737–744.
[13] Y. Li, Y. Song, J. Luo, Improving pairwise ranking for multi-label image classification, in: IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 3617–3625.
[14] S. Burkhardt, S. Kramer, Online multi-label dependency topic models for text classification, Machine Learning 107 (5) (2018) 859–886.
[15] Y. Sun, Y. Zhang, Z. Zhou, Multi-label learning with weak label, in: AAAI conference on Artificial Intelligence, 2010, pp. 593–598..

[16] Z. Chen, M. Chen, K. Weinberger, W. Zhang, Marginalized denoising for link prediction and multi-label learning, in: AAAI Conference on Artificial Intelligence, 2015, pp. 1707–1713..
[17] G. Yu, C. RangWala, H.and Domeniconi, G. Zhang, Z. Yu, Protein function prediction with incomplete annotations, IEEE/ACM Transactions on Computational Biology and Bioinformatics 11 (3) (2014) 579–591..
[18] D. Vasisht, M. Damianou, A. andVarma, A. Kapoor, Active learning for sparse bayesian multilabel classification, in: ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, 2014, pp. 472–481..
[19] H. Yu, P. Jain, P. Kar, I. Dhillon, Large-scale multi-label learning with missing labels, in: International Conference on Machine Learning, 2014, pp. 593–601..
[20] J. Bekker, J. Davis, Learning from positive and unlabeled data: a survey, Machine Learning 109 (4) (2020) 719–746.
[21] P. Konstantinos, V. Celine, Drug-target interaction prediction with tree-ensemble learning and output space reconstruction, BMC Bioinformatics 21 (49) (2020) 1–11.
[22] C. Elkan, K. Noto, Learning classifiers from only positive and unlabeled data, in: ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2008, pp. 213–220.
[23] D. Wang, L. Li, M. Zhang, Adaptive graph guided disambiguation for partial label learning, in: ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, 2019, pp. 83–91.
[24] L. Feng, B. An, Partial label learning with self-guided retraining, in: AAAI Conference on Artificial Intelligence, 2019, pp. 3542–3549..
[25] G. Lyu, S. Feng, T. Wang, C. Lang, Y. Li, Gm-pll: graph matching based partial label learning, IEEE Transactions on Knowledge and Data Engineering (2019) 1–14, https://doi.org/10.1109/TKDE.2019.2933837.
[26] G. Lyu, S. Feng, T. Wang, C. Lang, A self-paced regularization framework for partial label learning, IEEE Transactions on Cybernetics (2020) 1–13, https://doi.org/10.1109/TCYB.2020.2990908.
[27] M. Zhang, F. Yu, Solving the partial label learning problem: an instance-based approach, in: International Joint Conference on Artificial Intelligence, 2015, pp. 4048–4054..
[28] C. Gong, T. Liu, Y. Tang, J. Yang, J. Yang, D. Tao, A regularization approach for instance-based superset label learning, IEEE Transactions on Cybernetics 48 (3) (2017) 967–978.
[29] R. Jin, Z. Ghahramani, Learning with multiple labels, in: Advances in Neural Information Processing Systems, 2003, pp. 921–928.
[30] L. Liu, T. Dietterich, A conditional multinomial mixture model for superset label learning, in: Advances in Neural Information Processing Systems, 2012, pp. 548–556.
[31] M. Zhang, F. Yu, C. Tang, Disambiguation-free partial label learning, IEEE Transactions on Knowledge and Data Engineering 29 (10) (2017) 2155–2167.
[32] X. Wu, M. Zhang, Towards enabling binary decomposition for partial label learning., in: International Joint Conference on Artificial Intelligence, 2018, pp. 2868–2874..
[33] L. Feng, B. An, Leveraging latent label distributions for partial label learning, in: International Joint Conference on Artificial Intelligence, 2018, pp. 2107–2113..
[34] S. He, L. Feng, L. Li, Estimating latent relative labeling importances for multi-label learning, in: IEEE International Conference on Data Mining, 2018, pp. 1013–1018.
[35] N. Xu, J. Lv, X. Geng, Partial label learning via label enhancement, in: AAAI Conference on Artificial Intelligence, 2019, pp. 5557–5564..
[36] G. Lyu, S. Feng, Y. Li, Partial multi-label learning via probabilistic graph matching mechanism, in: ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2020, pp. 105–113.
[37] L. Feng, B. An, Partial multi label learning via multi-subspace representation, in: International Joint Conference on Artificial Intelligence, 2020, pp. 2612–2618..
[38] C. Wang, S. Yan, L. Zhang, Multi-label sparse coding for automatic image annotation, in: IEEE Conference on Computer Vision and Pattern Recognition, 2009, pp. 20–25.
[39] Y. Zhang, Z. Zhou, Multi label dimensionality reduction via dependence maximization, ACM Transactions on Knowledge Discovery from Data 4 (3) (2010) 1–21.
[40] G. Yu, H. RangWala, C. Domeniconi, G. Zhang, Protein function prediction using dependence maximization, in: European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases, 2013, pp. 574–589.
[41] G. Yu, J. Fu, J. Wu, H. Zhu, Predicting protein function prediction via semantic integration of multiple networks, IEEE/ACM Transactions on Computational Biology and Bioinformatics 13 (3) (2016) 220–232.
[42] Q. Tan, Y. Liu, X. Chen, G. Yu, Multi-label classification based on low rank representation for image annotation, Remote Sensing 9 (2) (2017) 109.
[43] Y. Yu, X. Yu, G.and Chen, Y. Ren, Semi-supervised multi-label linear discriminant analysis, in: International Conference on Neural Information Processing, 2017, pp. 688–698..
[44] L. Wu, R. Jin, A. Jain, Tag completion for image retrieval, IEEE Transactions on Pattern Analysis and Machine Intelligence 35 (3) (2013) 716–727.
[45] G. Zhu, S. Yan, Y. Ma, Image tag refinement towards low-rank, content-tag prior and error sparsity, in: ACM International Conference on Multimedia, 2010, pp. 461–470.
[46] M. Zhang, Z. Zhou, Ml-knn: A lazy learning approach to multi-label learning, Pattern Recognition 40 (7) (2007) 2038–2048.
[47] A. Elisseeff, J. Weston, A kernel method for multi-labelled classification, in: Advances in Neural Information Processing Systems, 2002, pp. 681–687.
[48] E. Gibaja, S. Ventura, A tutorial on multilabel learning, ACM Computing Surveys (CSUR) 47 (3) (2015) 52(1–39)..
[49] J. Demšar, Statistical comparisons of classifiers over multiple data sets, Journal of Machine Learning Research 7 (Jan) (2006) 1–30.