ORIGINAL ARTICLE

Constrained nonnegative matrix factorization-based semi-supervised multilabel learning

Dingguo Yu[1] · Bin Fu[2] · Guandong Xu[2] · Aihong Qin[1]

Abstract

In many multilabel learning applications, instances with labels being fully provided are scarce, while partially labelled data and unlabelled data are more common due to the expensive cost of manual labelling. However, most of existing models are based on the assumption that the fully labelled training data is sufficient. To deal with the partially labelled and unlabelled data effectively, we present a novel semi-supervised multilabel learning approach based on constrained non-negative matrix factorization in this paper. This approach assumes that if two instances are highly similar in terms of their features, they would also be similar in their associated labels set. Specifically, We first define three matrices to measure the similarity of each pair of instances in two different ways. Then, the optimal assignation of labels to the unlabelled instance is determined by minimizing the differentiation between these two similarity sets via a non-negative matrix factorization process. We also present a threshold learning algorithm to determine the classification threshold for each label in our proposed approach. Extensive experiment is conducted on various datasets, and the results demonstrate that our method show significantly better performance than other state-of-the-art approaches. It is especially suitable for the situations with a smaller size of labelled training data, or subset of the training data are partially labelled.

Keywords Semi-supervised learning · Nonnegative matrix factorization (NMF) · Multilabel learning · Weak label

1 Introduction

In traditional machine learning, an instance is usually assumed to have only one class label. However, in real applications, an instance usually consists of multiple concepts simultaneously. For instance, in text categorization, a news report could cover several topics; while in scene classification, an image could be related to some scenarios, to name a few. Accordingly, learning models that could predict multiple labels simultaneously for an instance is called multilabel learning [22]. In multilabel learning, each instance is also represented by a vector of features as in traditional single-label classification, while associated with multiple labels instead of a single label [30]. Nowadays, there have been extensive applications of multilabel learning, such as text categorization [20], gene function analysis [5] and image or video annotation [17], et al.

Over the last decade, a variety of multilabel learning models have been proposed, such as Binary Relevance [3], Classifier Chains [19], ML-kNN [29], ML-DT [5], Rank-SVM [6], MIAL [25] and so on. More details on these algorithms are described in the related work section. Generally, most of these approaches assume that for every instance, all its labels have been provided, and the training dataset is sufficiently to learn a stable classifier. However, in real applications, the expensive cost on manual labelling process possibly makes a large size of full labelled training data unfeasible, while obtaining partial labeled data or a small size of full labelled data is easy relatively. Furthermore, some researchers found that a great amount of unlabelled data used together with labelled data for training, can be generated a considerable improvement on prediction accuracy [2]. In this case, the semi-supervised multilabel learning would be of a huge practical values, and many semi-supervised learning algorithms have been proposed and applied [1, 31].

In addition, some multilabel learning applications only have a small size of training dataset but with a great number

---

✉ Dingguo Yu
yudg@zjicm.edu.cn

1 School of New Media, Zhejiang University of Media and Communications, Hangzhou 310018, China

2 Advanced Analytics Institute, University of Technology, Sydney, NSW 2007, Australia

of labels. Consequently, the training dataset related to each label is not sufficient for training a stable classifier. To deal with this issue, Liu et al. present a semi-supervised multilabel learning approach [16], which is based on the key assumption that two instances tend to have large overlap in their assigned class memberships if they share high similarity in their input patterns. This approach explores the unlabelled data and the class correlation simultaneously, and addresses the problem arising from a small size of training data, but it still needs the training instance with full labels. However, as a matter of fact, it is very difficult to get full labels for each instance, especially when the size of labels in a multilabel domain is huge. Usually, only partial labels are available. To solve this issue, we propose a novel approach for semi-supervised multilabel learning based on the constrained non-negative matrix factorization (NMF) model. We name it as NMFSML in this paper. Our approach is also based on the assumption described above, but the difference is that we apply NMF method to deal with the optimization task. The advantage of NMFSML is that it does not need the training instance with full labels and it is more efficient. Specifically, we first define three matrices to measure the similarity of all possible pairs of instances in two different ways, i.e., one based on their features (input pattern), and another based on the label memberships (output pattern). Utilizing the nonnegative matrix factorization technology, we then learn the optimal labels for unlabelled instances through minimizing the differentiation between these two similarity sets.

The main contributions of this paper are summarized as following:

1. We present a novel semi-supervised multilabel learning framework based on constrained NMF (NMFSML), explore the iterative updating rules, and implement the optimization process. Compared to existing approaches for multilabel learning, NMFSML is more efficient for such applications with a small size of training data and even allow including data with partial labels, especially for the learning situation with a great amount of labels.
2. We propose a classification threshold learning algorithm combine with our proposed leaning approach, which can solve the threshold setting issue within multilabel classification algorithms.
3. We conduct extensive experiment on various datasets from different domains, and compare our approach with other approaches. The results demonstrate that our approach performs significantly better than the other approaches.

The remaining of this paper is outlined as follows. Section 2 introduces related works regarding multilabel learning. Section 3 details the proposed novel approach for

semi-supervised multilabel learning based on constrained non-negative matrix factorization, including the definition of semi-supervised multilabel learning, the iterative update rules, the optimization process of NMFSML, and the classification threshold learning algorithm. Section 4 reports the experiment design and results analysis. Section 5 concludes our work.

2 Related works

In the last decade, a diversity of multilabel learning approaches have been presented and successfully applied in various areas. These approaches can be roughly divided into two categories: problem transformation approaches and algorithm adaptation approaches.

For problem transformation approaches, the basic idea is to transform the multilabel learning problem into other learning problem. A typical approach of this type is Binary Relevance, which decomposes the multilabel learning into a set of independent binary classifications, each binary classification corresponding respectively to a class [3]. The drawback of this method is that it ignores the correlation between different labels, which could provide useful information to assist label prediction. To address this issue, many methods that incorporate the class correlation into multilabel learning have been developed in recent years, including Classifier Chains [19], IBLR-ML+ [4], LEAD [28], CDN-LR [12], SELD [9] and so on. Additionally, some approaches such as Calibrated Label Ranking algorithm [10] transform it into a problem of label ranking, and some approaches convert the multilabel learning problem into a set of multiclass classification problems.

Approaches that based on algorithm adaptation solve the multilabel learning problem by developing the common learning algorithm to directly handle the multilabel data, typical approaches include ML-kNN algorithm which adapts lazy learning techniques [29], ML-DT approach which alters decision tree model [5], Rank-SVM algorithm which modifies kernel techniques [6], and CML approach which improves information theoretic techniques [11], et al.

Both of the problem transformation approaches and algorithm adaptation approaches are based on the assumption, which the training dataset is enough for learning stable classifiers and all labels of every instance are provided. However, this assumption is not true in some real applications. Furthermore, these approaches ignore that there are usually a huge amount of unlabelled instances, when used with the labelled instances, could boost the learning performance significantly.

Nonnegative matrix factorization (NMF) [14] is a popular method for finding parts-based representations of nonnegative data. Due to the dimension reduction effectiveness of

NMF [13, 18], it has been successfully applied for multilabel classification such as PLST [21], MLC-BMaD [26], CNMF [16] and so on. However, the original NMF is an unsupervised learning algorithm, it cannot make full use of labeled data to improve learning performance, so many researchers develop the usage of NMF to a semi-supervised manner [8, 24, 27], and so on. In this paper, we present a novel approach for semi-supervised multilabel learning based on constrain NMF, our proposed approach is effective for this learning situation with a small size of training dataset but with a great number of labels. Most importantly, it does not require each training instance with all their labels provided.

3 Our Proposed NMFSML Approach

3.1 Semi-supervised multiLabel learning

We formally define the semi-supervised multilabel learning problem as follows. Denote $X = \{x_1, x_2, \ldots, x_{l-1}, x_l, x_{l+1}, \cdots, x_n\} \subset R^m$ as the entire dataset, where n is the total number of instances, including the labelled instances and the unlabelled instances, the labelled instances includes fully labelled data and partially labelled data. Without loss of generality, we let the first l instances be the labelled instances. Let $C = \{c_1, c_2, \ldots, c_q\}$ be a set of labels, where q is the total number of labels. For the labeled instances, their label information can be presented in the matrix $Y = [y_{ij}]_{l \times q}$ as the *label matrix*, where $y_{ij} = 1$ indicates label c_j is instance $x_i's$ relevant or true label, $y_{ij} = 0$ indicates label c_j is not instance $x_i's$ relevant or not true label, and $y_{ij} = -1$ indicates that it is unknown whether the label c_j is instance $x_i's$ relevant or true label. The task of semi-supervised multilabel learning is to learn a function $h : X \to 2^C$ from the dataset X, which is used to predict the unknown labels for the test data.

3.2 The framework of NMFSML

In this section, we will detail our proposed novel approach for semi-supervised multilabel learning. Our work is also based on the assumption which if two instances are highly similar in terms of their features, they would also be similar in their associated labels set. We measure the similarity between two instances in two ways respectively, one is called input pattern similarity that is based on the correlation between their features, and another is called output pattern similarity that is based on the correlation between their label memberships. According to the assumption, for a pair of instances, their input pattern similarity and output pattern similarity should be similar. Hence we can predict the optimal assignment of label memberships to an unlabelled instance through minimizing the differentiation

between these two similarity sets, and this optimization task can be transformed into a non-negative matrix factorization problem.

In order to calculate the input pattern similarity and output pattern similarity, similar to Liu et al. proposed approach [16], we also introduce three matrices, but the calculating method of these matrices are not same, the difference will be given in the following section. (1) Normalized matrix $A = [a_{ij}]_{n \times n}$ as the *input pattern similarity matrix*, where element $a_{ij} \geq 0$ represents the similarity measurement between the two instances x_i and x_j based on their input patterns, it can be computed from their features using Euclidean distance or cosine similarity in the m-dimensional space; (2) Matrix $B = [b_{kl}]_{q \times q}$ as the *label similarity matrix*, where element $b_{kl} \geq 0$ represents the similarity measurement between two labels c_k and c_l; and (3) Matrix $T = [t_{ik}]_{n \times q}$ as the *label confidence matrix*, where element $t_{ik} \geq 0$ is the confidence score of assigning label c_k to instance x_i. For considering the scale mismatch between the two similarities, it need to normalize these matrices after it has been computed.

In Liu et al. proposed approach, it need first compute the input pattern similarity matrix A and label similarity matrix B, and then use them and the assigned class labels information to iterative compute the label confidence matrix T, so it still needs the training instances with full labels. While in our approach, the label confidence matrix T is computed via constrained non-negative matrix factorization technology by using the input pattern similarity matrix A and label matrix Y, without direct calculation of the label similarity matrix B, so it can address the weak label problem.

According to the matrices defined above, for any two instances x_i and x_j, their label confidence vector is $T_i = (t_{i1}, t_{i2}, \cdots, t_{iq})$ and $T_j = (t_{j1}, t_{j2}, \cdots, t_{jq})$ respectively. Therefore, the measurement of output pattern similarity between instances x_i and x_j can be represented as $T_i B T_j^T$, while their measurement of input pattern similarity is a_{ij}. Following the assumption above, we expect that these two measurements of similarity closer, denote as $a_{ij} \approx T_i B T_j^T$. It can be obtained the optimization problem as follows:

$$\arg \min_{T, B \geq 0} \sum_{i=1}^{n} \sum_{j=1}^{n} (a_{ij} - T_i B T_j^T)^2 \tag{1}$$

Furthermore, the formula in Eq. (1) would be reformulated as the following NMF problem:

$$\arg \min_{T, H \geq 0} \|A - TH\|_F^2 \tag{2}$$

where $\|\cdot\|_F$ stands for the Frobenius norm, and $H = BT^T$.

In the semi-supervised multilabel learning, the training instances include labelled instances and unlabelled instances. For each labelled instance, their estimated label confidence score t_{ik} should be equal with the associated label value y_{ik}.

Therefore, we refer to the associated label information as constraint on NMF process, and the optimization problem is depicted as follows:

$$arg \min_{T,H \geq 0} \{\|A - TH\|_F^2 + \alpha(\|T\|_F^2 + \|H\|_F^2)\}$$

$$s.t. \quad t_{ik} = y_{ik} \quad (3)$$
$$(y_{ik} \neq -1; i = 1, 2, \ldots, l; k = 1, 2, \ldots, q)$$

where α is a smooth regularization parameter and $\alpha(\|T\|_F^2 + \|H\|_F^2)$ is a regularization term serving as smoothness constraint to improve the robustness of the learned model.

3.3 Optimization algorithm

In this section, we explore an iterative updating algorithm which is used to get the local optima of Eq. (3). Based on the matrix property $Tr(AB) = Tr(BA)$, the objective function can be written as:

$$O_F = \|A - TH\|_F^2 + \alpha(\|T\|_F^2 + \|H\|_F^2)$$
$$= Tr((A - TH)(A - TH)^T) + \alpha(\|T\|_F^2 + \|H\|_F^2)$$
$$= Tr(AA^T) - 2Tr(AH^T T^T) + Tr(THH^T T^T) + \alpha(\|T\|_F^2 + \|H\|_F^2)$$

Denote φ and ϕ as the Lagrange multiplier to constraints $t_{ij} \geq 0$ and $h_{ij} \geq 0$, respectively. The Lagrange function L is

$$L = O_F + Tr(\varphi T^T) + Tr(\phi H^T)$$

We get the following equations by requesting the derivatives of L with respect to T and H respectively.

$$\frac{\partial L}{\partial T} = -2AH^T + 2THH^T + 2\alpha T + \varphi = 0$$
$$\frac{\partial L}{\partial H} = -2T^T A + 2T^T TH + 2\alpha H + \phi = 0$$

Under the Kuhn-Tucker condition $\varphi_{ij} t_{ij} = 0$ and $\phi_{ij} h_{ij} = 0$, we can obtain the equations of y_{ij} and h_{ij} as follows:

$$(AH^T)_{ij} t_{ij} - (THH^T + \alpha T)_{ij} t_{ij} = 0$$
$$(T^T A)_{ij} h_{ij} - (T^T TH + \alpha H)_{ij} h_{ij} = 0$$

So, the above equations can lead to the updating rules:

$$t_{ij} \leftarrow t_{ij} \frac{(AH^T)_{ij}}{(THH^T + \alpha T)_{ij}} \quad (4)$$

$$h_{ij} \leftarrow h_{ij} \frac{(T^T A)_{ij}}{(T^T TH + \alpha H)_{ij}} \quad (5)$$

By using the auxiliary function method, we can prove the convergence of this updating rules [15, 27].

To sum up, the procedure to solve the optimization problem proposed above could be formulated as in Algorithm 1.

The original NMF is an unsupervised learning algorithm, we extent the standard NMF to a semi-supervised learning algorithm by taking the label information as a hard constraint in this paper. In the semi-supervised multilabel learning, the training instances include labelled instances and unlabelled instances, for some labelled instances, only a "partial" label set is available probably (weak label problem). As the definition of label matrix in Sect. 3.1, where $y_{ij=-1}$ indicates that it is unknown whether the label c_j is instance x_i's relevant or true label. In our proposed method, we refer to the associated label information as constraint on NMF process, for each labelled instance, their estimated label confidence score t_{ij} should be equal with the associated label value y_{ij}. So in Algorithm 1, where $i \leq l$ and $y_{ij} \neq -1$, we let $t_{ij} = y_{ij}$, otherwise calculate t_{ij} using NMF method. Then our proposed method can address the weak label problem.

Algorithm 1: Label Constrained NMF Algorithm

Input: *Input pattern similarity matrix A; Label matrix Y;* Parameter α

Output: *Matrices T and H*

Steps:
(1) Initialize T and H by random positive values
(2) **Repeat**
(3) if $(i \leq l)$ && $(y_{ij} \neq -1)$
$$t_{ij} = y_{ij}$$
(4) else
$$t_{ij} = t_{ij} \frac{(AH^T)_{ij}}{(THH^T + \alpha T)_{ij}}$$
$$h_{ij} = h_{ij} \frac{(T^T A)_{ij}}{(T^T TH + \alpha H)_{ij}}$$
(5) **Until** Convergence criterion is met
(6) Output T and H

3.4 Label predicting

For a new instance x_u, we denote its input pattern similarity as a normalized vector $A_u = (a_{u1}, a_{u2}, \cdots, a_{un})$, wherein element a_{uj} $(1 \leq j \leq n)$ represents the input pattern similarity measurement between the new instance x_u and the training instance x_j. Then the label confidence vector $T_u = (t_{u1}, t_{u2}, \cdots, t_{uq})$ can be computed by the following equation.

$$T_u = A_u H^{-1} \quad (6)$$

where matrix H is the output of NMF, H^{-1} is the generalized inverse matrix of matrix H.

After that, we would predict the optimal assignation of labels to the new instance x_u as follows:

$$y_{uk} = \begin{cases} 1, & t_{uk} \geq \varepsilon_k \\ 0, & t_{uk} < \varepsilon_k \end{cases} \quad (7)$$

where ε_k is the classification threshold of class label c_k (the calculation method of classification threshold will be presented in the next section), t_{uk} is the confidence score of predicting the k-th label c_k to the new instance x_u. While $y_{uk} = 1$ indicates the class label c_k is $x_i's$ relevant or true label, $y_{uk} = 0$ indicates class label c_k is not $x_i's$ relevant or not true label.

In summary, the steps of predicting the optimal assignment of class memberships to the new instance could be formulated as in Algorithm 2.

Algorithm 2: Label predicting algorithm

Input: Training dataset X; matrix H; new instance x_u; classification threshold vector ε.

Output: Label vector Y_u;

Steps:

(1) Computing the input pattern similarity normalized vector A_u

(2) Computing the label confidence vector by Eq. (6)

(3) Predicting the optimal assignment of class memberships using Eq. (7)

(4) Output Y_u

3.5 Classification threshold learning

The classification threshold determination is an important issue in multilabel learning, the researchers conducted extensive research on threshold determination problem, such as Read et al. put forward the threshold calibration method [19], Fan et al. propose the SCutFBR algorithm [7] and so on. In our early research experiments, we have observed that some threshold calibration methods that are independent of the specific multilabel learning algorithm do not achieve efficient classification outcomes, such as that it is very inefficient by simply using an arbitrary threshold like 0.5. In the next section, we will propose the threshold calibration method which in agreement with our proposed multilabel learning method NMFSML based on the training dataset.

For one class label $c_k \in C$ $(1 \leq k \leq q)$, we denote the accept threshold, the reject threshold and the classification threshold as p_k, r_k and ε_k respectively, and the classification threshold vector as $\varepsilon = (\varepsilon_1, \varepsilon_2, \cdots, \varepsilon_q)$. Firstly, we calculate the confidence score vector $T_i = (t_{i1}, t_{i2}, \cdots, t_{iq})$ for each labelled instance $x_i \in X$ $(1 \leq i \leq l)$ by using Eq. (6), then compute the p_k, r_k and ε_k as follows:

$$p_k = avg\{t_{ik} | y_{ik} = 1, 1 \leq i \leq l\} \tag{8}$$

$$r_k = avg\{t_{ik} | y_{ik} = 0, 1 \leq i \leq l\} \tag{9}$$

$$\varepsilon_k = avg\{p_k, r_k\} \tag{10}$$

where $avg\{\cdot\}$ stands for the averaging function.

Table 1 Datasets used in experiments

Name	Domain	$|D|$	$F(D)$	$L(D)$	$LD(D)$
Scene	Image	2407	294	6	1.074
Birds	Audio	645	260	19	1.014
Yeast	Biology	2417	103	14	4.237
Cal500	Music	502	68	174	26.044

4 Experiments

In this section, we report the experiments conducted on several typical datasets to evaluate our proposed approach, we compare it with other state-of-the-art methods and evaluate their performance.

4.1 Dataset and evaluation metric

We carry out experiments on four typical datasets that described in Table 1. These datasets come from a variety of domains including image, music, biology and audio, which are extensively used for evaluating multilabel learning methods, more detailed description can be obtained from the website[1]. For one dataset D, we denote $|D|$, $F(D)$, $L(D)$ and $LD(D)$ as the number of instances, number of features, number of possible labels, and average number of labels per example in dataset respectively.

We use the standard information retrieval metric F1-measure to evaluate the effectiveness and efficiency of our proposed approach. F1-measure value is more precise to evaluate the performance of classifiers, because it is a combination value with summarizing of both precision and recall.

4.2 Baselines and experimental setup

We compare our model (NMFSML) with four baselines approaches: (1) **Classifier Chains (CC)**, this algorithm transforms a multilabel learning task into a binary classification tasks chain, and each binary classifier is depend on the previous ones [19]; (2) **IBLR-ML+**, which unifies instance-based learning and logistic regression, and combines model-based and similarity-based inference for multilabel classification [4]; (3) **Calibrated Label Ranking (CLR)**, which transforms a multilabel learning task into a label ranking task, where ranking among labels is fulfilled by pairwise comparison [10]; (4) **CNMF**, which is a semi-supervised multilabel learning by constrained non-negative matrix factorization [16]. The fourth method is similar to our approach. For IBLR-ML+, we let the nearest neighbors
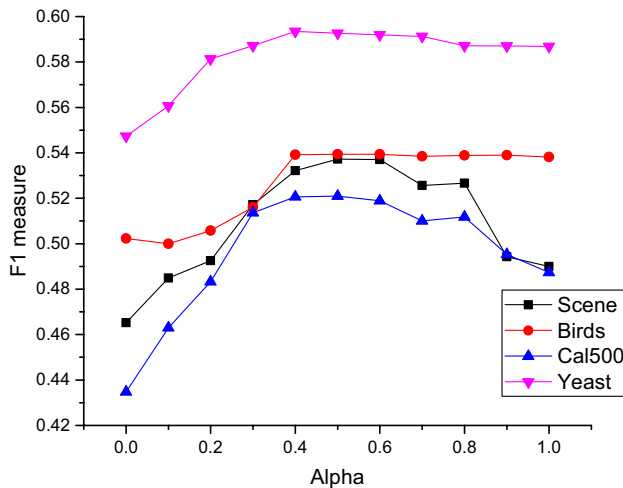
---

[1] http://mulan.sourceforge.net/datasets-mlc.html

Fig. 1 Impact of the Parameter α

as 10. For CNMF, the input pattern similarity matrix A is computed as the cosine similarity between their features, the class similarity matrix B is implemented through computing the pairwise class similarity based on their vector representation (each class c is represented as a binary vector whose elements are set to be one if the corresponding training example belong to the class c and zero otherwise). The CC, IBLR-ML+ and CLR are implemented in Mulan [23], which is an open source framework designed for multilabel learning.

In order to improve the quality of the experimental results, according to the convergence of iterative rules, we set the convergence criteria as $(|(t_{ij})_{n+1} - (t_{ij})_n| \le 10^{-4})$ $\&\&(|(h_{ij})_{n+1} - (h_{ij})_n| \le 10^{-4})$ in Algorithm 1, where n is the iterative computation times.

In addition, for Eq. (3), the effect of parameter α need to be further explored, α is to control the smoothness and robustness of the learned model. To explore the parametrical stability of our approach, we evaluate the performance (F1-measure) of NMFSML on the all datasets under a series of varying parameter settings. We randomly select 50% data points from the dataset as the labelled data, and then increase α gradually from 0 to 1 with a step size of 0.1, the experimental results is plotted in Fig. 1. The results suggest that the proposed framework NMFSML can achieve a relatively good performance when the parameter α is in the range [0.4, 0.6], we choose $\alpha = 0.5$.

4.3 Experiment design and results analysis

To comprehensively evaluate the performance of our approach NMFSML, we design two experiments as follows. All programs are run on datasets using five-fold cross validation, and the result shown in the following figures is the average value.

Experiment 1: Compare NMFSML with other approaches. We randomly select a portion of instances from the dataset as the labelled set, select 20% instances as the testing data, and the remaining part of instances will be used as the unlabelled data (for NMLSML and CNMF). The portion of labeled data gradually increases from 5% to 70% with a step size of 5% and the results as shown in Fig. 2.

Experiment 2: The performance based on partially labelled data. We randomly select 50% instances from the dataset as the labelled data, and the remaining part of instances are used as the unlabelled data and the testing set. We randomly change a portion of known labels to unknown labels in the labelled data, then the full labels data will be changed to partial labels data. The portion of the partial labels data gradually increases from 0% to 40% with a step size of 5% and the result as shown in Fig. 3. This experiment is only to show the effectiveness of our approach to solve the weak label problem, and we will not compare it with other methods and discuss the performance due to space limitations.

According to the comparative analysis based on the experiment results, we can get the findings as follows. From Fig. 2, we observe that, these methods are all showing the increasing trend in terms of F1 measure with the increase of labelled data, but NMFSML and CNMF make more significant improvement compared to another three baseline approaches, especially on the learning scenario with a small size of dataset for training (at the lower percentage of labelled data) or with a large number of labels. Because in these learning scenarios where the training dataset associated to each label is insufficient, NMFSML and CNMF can utilize the unlabelled data to assist label prediction, but CC, IBLR-ML+ and CLR cannot. Additionally, from Fig. 3, we can see that, the NMFSML shows a trend of decreasing F1-measure when the percentage of partially labelled data increases, but the rate of F1-measure decline is far less than the increasing rate of partially labelled data. Therefore, the NMFSML can solve the weak label problem effectively, but CNMF cannot.

5 Conclusions

In this paper, we propose a novel framework for semi-supervised multilabel learning based on constrained NMF, and present a threshold learning algorithm which can determine the classification threshold for each label within our NMFSML model. The advantage of our proposed method is that it not only utilize the relation between different labels, but also exploit the unlabelled data simultaneously for learning, and the learned representations can
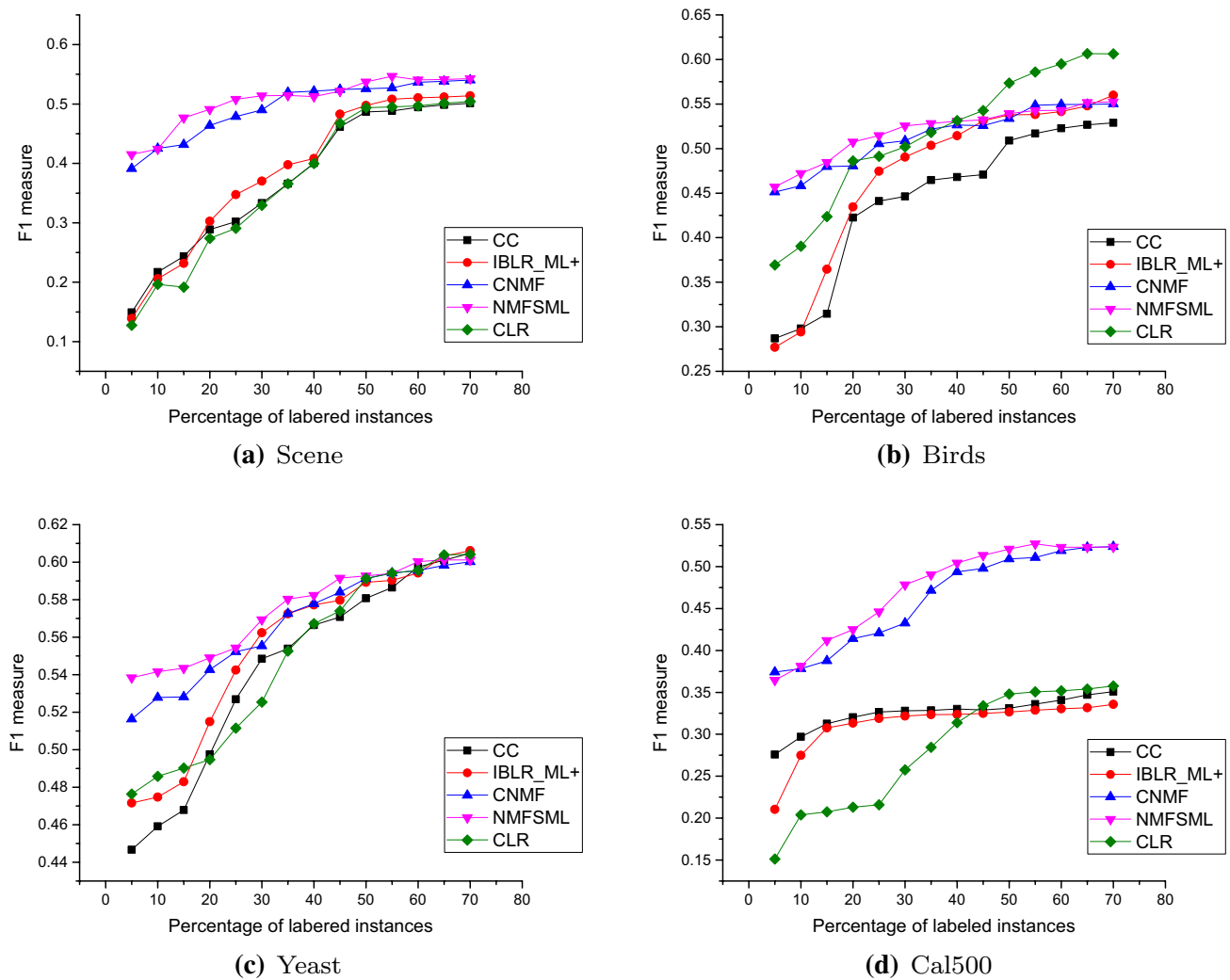
Fig. 2 Compare NMFSML with other approaches. **a** Scene. **b** Birds. **c** Yeast. **d** Cal500
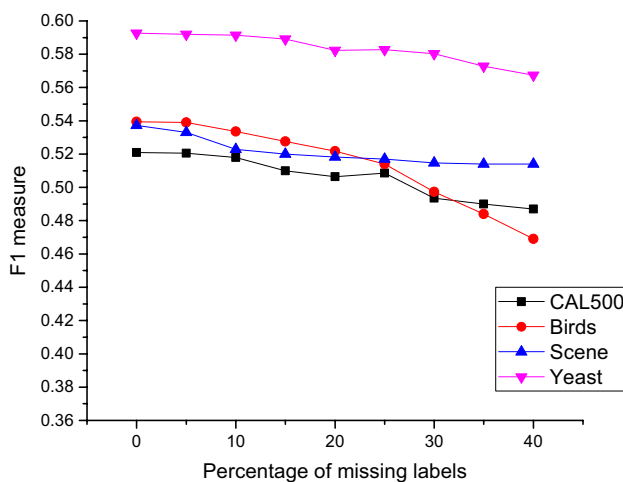

Fig. 3 The performance based on partial labels data

have more discriminating power by using the information provided by a few labelled instances and a large number of unlabelled instances. We have conducted extensive experiment on a board rage of datesets, and the results indicate that our method outperforms other state-of-the-art multilabel learning methods significantly. The NMFSML is especially suitable for the learning scenarios with a small size of labelled data for training, or the training data include partially labelled data.

References

1. Ashfaq RAR, Wang XZ, Huang JZ, Abbas H, He YL (2017) Fuzziness based semi-supervised learning approach for intrusion detection system. Inf Sci 378:484–497

2. Belkin M, Niyogi P, Sindhwani V (2006) Manifold regularization: a geometric framework for learning from labeled and unlabeled examples. J Mach Learn Res 7:2399–2434

3. Boutell MR, Luo J, Shen X, Brown CM (2004) Learning multi-label scene classification. Pattern Recogn 37(9):1757–1771

4. Cheng W, Hüllermeier E (2009) Combining instance-based learning and logistic regression for multilabel classification. Mach Learn 76(2–3):211–225

5. Clare A, King RD (2001) Knowledge discovery in multi-label phenotype data. In: European Conference on Principles of Data Mining and Knowledge Discovery, Springer, New York, pp 42–53

6. Elisseeff A, Weston J (2001) A kernel method for multi-labelled classification. In: Proceedings of the 14th international conference on neural information processing systems: natural and synthetic. MIT Press, pp 681–687

7. Fan RE, Lin CJ (2007) A study on threshold selection for multi-label classification. Department of Computer Science, National Taiwan University, pp 1–23

8. Feng X, Jiao Y, Lv C, Zhou D (2016) Label consistent semi-supervised non-negative matrix factorization for maintenance activities identification. Eng Appl Artif Intell 52:161–167

9. Fu B, Xu G, Wang Z, Cao L (2013) Leveraging supervised label dependency propagation for multi-label learning. 2013 IEEE 13th international conference on data mining (ICDM), IEEE, pp 1061–1066

10. Fürnkranz J, Hüllermeier E, Mencía EL, Brinker K (2008) Multilabel classification via calibrated label ranking. Mach Learn 73(2):133–153

11. Ghamrawi N, McCallum A (2005) Collective multi-label classification. In: Proceedings of the 14th ACM international conference on information and knowledge management. ACM, pp 195–200

12. Guo Y, Gu S (2011) Multi-label classification using conditional dependency networks. In: IJCAI proceedings–international joint conference on artificial intelligence, vol 22, no 1, p 1300

13. Kimura K, Kudo M, Sun L (2015) Dimension reduction using nonnegative matrix tri-factorization in multi-label classification. In: Proceedings of the international conference on parallel and distributed processing techniques and applications (PDPTA), p 250

14. Lee DD, Seung HS (1999) Learning the parts of objects by non-negative matrix factorization. Nature 401(6755):788–791

15. Liu H, Wu Z, Li X, Cai D, Huang TS (2012) Constrained nonnegative matrix factorization for image representation. IEEE Trans Pattern Anal Mach Intell 34(7):1299–1311

16. Liu Y, Jin R, Yang L (2006) Semi-supervised multi-label learning by constrained non-negative matrix factorization. In: Proceedings of the National Conference on Artificial Intelligence, vol 21, Menlo Park, Cambridge, London; AAAI Press; MIT Press

17. Qi GJ, Hua XS, Rui Y, Tang J, Mei T, Zhang HJ (2007) Correlative multi-label video annotation. In: Proceedings of the 15th international conference on Multimedia, ACM, pp 17–26

18. Qian B, Davidson I (2010) Semi-supervised dimension reduction for multi-label classification. AAAI 10:569–574

19. Read J, Pfahringer B, Holmes G, Frank E (2011) Classifier chains for multi-label classification. Mach Learn 85(3):333–359

20. Schapire RE, Singer Y (2000) Boostexter: A boosting-based system for text categorization. Mach Learn 39(2):135–168

21. Tai F, Lin HT (2012) Multilabel classification with principal label space transformation. Neural Comput 24(9):2508–2542

22. Tsoumakas G, Katakis I, Vlahavas I (2009) Mining multi-label data. In: Maimon O, Rokach L (eds) Data mining and knowledge discovery handbook. Springer, Boston, MA, pp 667–685

23. Tsoumakas G, Spyromitros-Xioufis E, Vilcek J, Vlahavas I (2011) Mulan: a java library for multi-label learning. J Mach Learn Res 12:2411–2414

24. Wang D, Gao X, Wang X (2016) Semi-supervised nonnegative matrix factorization via constraint propagation. IEEE Trans Cybern 46(1):233–244

25. Wang R, Wang XZ, Kwong S, Xu C (2017) Incorporating diversity and informativeness in multiple-instance active learning. IEEE Trans Fuzzy Syst 25(6):1460–1475

26. Wicker J, Pfahringer B, Kramer S (2012) Multi-label classification using boolean matrix decomposition. In: Proceedings of the 27th annual ACM symposium on applied computing. ACM, pp 179–186

27. Yu D, Chen N, Jiang F, Fu B, Qin A (2017) Constrained nmf-based semi-supervised learning for social media spammer detection. Knowl-Based Syst 125:64–73

28. Zhang ML, Zhang K (2010) Multi-label learning by exploiting label dependency. In: Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining, ACM, pp 999–1008

29. Zhang ML, Zhou ZH (2007) Ml-knn: A lazy learning approach to multi-label learning. Pattern Recogn 40(7):2038–2048

30. Zhang ML, Zhou ZH (2014) A review on multi-label learning algorithms. IEEE Trans Knowl Data Eng 26(8):1819–1837

31. Zhu H, Wang X (2017) A cost-sensitive semi-supervised learning model based on uncertainty. Neurocomputing 251:106–114