Matrix Factorization for Identifying Noisy Labels of Multi-label Instances

Xia Chen[1][0000−0002−8223−5641], Guoxian Yu[1]✉[0000−0002−1667−6705], Carlotta Domeniconi[2], Jun Wang[1][0000−0002−5890−0365], and Zili Zhang[1]

[1] College of Computer and Information Sciences, Southwest University, Chongqing, 400715, China {xchen,gxyu,kingjun,zhangzl}@swu.edu.cn
[2] Department of Computer Science, George Mason University, Fairfax VA, 22030, USA carlotta@cs.gmu.edu

Abstract. Current effort on multi-label learning generally assumes that the given labels are noise-free. However, obtaining noise-free labels is quite difficult and often impractical. In this paper, we study how to identify a subset of relevant labels from a set of candidate ones given as annotations to instances, and introduce a matrix factorization based method called *MF-INL*. It first decomposes the original instance-label association matrix into two low-rank matrices using nonnegative matrix factorization with feature-based and label-based constraints to retain the geometric structure of instances and label correlations. MF-INL then reconstructs the association matrix using the product of the decomposed matrices, and identifies associations with the lowest confidence as noisy associations. An empirical study on real-world multi-label datasets with injected noisy labels shows that MF-INL can identify noisy labels more accurately than other related solutions and is robust to input parameters. We empirically demonstrate that both feature-based and label-based constraints contribute to boosting the performance of MF-INL.

Keywords: Multi-label learning · Noisy labels identification · Low-rank matrix factorization.

1 Introduction

Multi-label classification models the scenario where each instance is associated with a set of labels, and its goal is to find a set of relevant labels for unlabeled instances [36,6]. Multi-label classification has attracted ever-increasing interest in the context of text classification [18], automatic image annotation [23], and protein function prediction [28], among other applications. Currently, multi-label learning methods mainly focus on how to assign a set of appropriate labels to unlabeled instances [35], how to replenish missing labels for incompletely labeled instances [19], and how to make use of interrelationships of labels [33]. Most of the aforementioned methods assume that the assigned labels are correct. However, things may go athwart in practice, and the collected label set of observed instances may include *noisy* (or not applicable) labels. This is because the labels of multi-label instances are collected by human annotators with wide-ranging levels of expertise and different techniques [29,14].

Despite the progress achieved in multi-label learning, the problem of identifying noisy labels in multi-label instances, to the best of our knowledge, is *seldom* studied. Its goal is the selection of a set of appropriate labels for a multi-label instance, by removing from the given collection those that do not apply. As such, the problem is more challenging and different from partial-label learning [4,34], which identifies one label from a set of candidate labels of an instance, disregarding their interrelationship.

In this paper, we propose a matrix factorization based approach called MF-INL to identify noisy labels of multi-label instances. MF-INL first factorizes the instance-label association matrix into two low-rank matrices via graph regularized Nonnegative Matrix Factorization (NMF) [13,2]. Particularly, MF-INL takes advantage of the geometric structure among instances and correlations between labels to define two graphs, and thus enforces the factorized matrices to be consistent with both the geometric structure and label correlations. After that, MF-INL reconstructs the association matrix using the product of the two factorized matrices, and considers the reconstructed associations with low entry values as noisy labels of instances. Experimental results on publicly available multi-label datasets show that MF-INL can identify noisy labels of multi-label instances more accurately than other related methods [31,34,3,27,15].

2 Related work

Partial-label learning studies the scenario in which an instance is associated with a set of candidate labels among which only one is valid [4,26], and can be viewed as a special case of multi-label partial-label learning, where we force each instance to be annotated with one label only. Most partial-label learning methods combine the ground-truth label identification and classifier training on the over-labeled instances [32]. Some methods treat equally all candidate labels and make prediction by averaging the outputs of all candidate labels [4] [8]. Other methods assume a parametric model and consider the ground-truth labels as latent variables, which are iteratively refined to disambiguate candidate labels [16,21]. More recent methods follow two stages: they first directly disambiguate the candidate labels, and then perform classification on the disambiguated labels of instances [31,34]. All the aforementioned partial-label learning methods assume that each instance is associated with exactly one ground-truth label. But in many data mining application domains, instances are naturally associated with multiple labels.

LSDR methods have been proposed for multi-label learning to handle large label space [20]. CPLST [3] simultaneously considers both the instance-label association information and the instance-feature information to minimize the upper bound of popular Hamming loss, and consequently to seek a latent label space. FaIE [15] jointly maximizes the recoverability of the original label space from the latent space, and the predictability of the latent space from the feature space, to simultaneously learn the coding and decoding matrices, which are used to compress and recover the label space, respectively. All these LSDR based methods aim to find an optimal low-dimensional subspace with respect to the original label space, and to perform multi-label learning in the subspace to improve performance by removing irrelevant, redundant, or noisy information (i.e., noisy labels of instances) [3]. In addition, some other recent methods also show the contribution of label embedding for multi-label classification [24,25].

Motivated by the robustness to noise of low-rank matrix approximation [7,17], we introduce a matrix factorization based approach MF-INL to identify noisy labels of multi-label instances. The empirical study shows that MF-INL can identify noisy labels more accurately than competitive algorithms.

3 Proposed method

Let $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_n] \in \mathbb{R}^{d \times n}$ be n multi-label instances in the d-dimensional feature space. $\mathbf{Y} \in \mathbb{R}^{q \times n}$ is the known instance-label association matrix. $\mathbf{Y}_{ci} \in \{0, 1\}$, where $\mathbf{Y}_{ci} = 1$ if $\mathbf{x}_i$ is associated with the c-th label, $\mathbf{Y}_{ci} = 0$ otherwise. The goal of MF-INL is to identify noisy associations in $\mathbf{Y}$.

The low-rank approximation of a *noisy* matrix is robust to noise [12,17]. Thus, in this paper, we consider to seek the low-rank approximation of the original instance-label association matrix to identify noisy labels. Particularly, we advocate to decompose the instance-label association matrix $\mathbf{Y}$ into two low-rank matrices by NMF, which is a widely used low-rank matrix decomposition method. Then, we can use the product of two low-rank matrices to approximate the original matrix. However, this decomposition does not consider the geometric structure among instances and correlations among labels, both of which should be leveraged to guide the decomposition. Our proposed method MF-INL addresses this issue by integrating nonnegative matrix factorization with feature-based and label-based constraints. MF-INL minimizes the objective function as follows:

$$\psi(\mathbf{U}, \mathbf{V}) = \|\mathbf{Y} - \mathbf{U}\mathbf{V}^T\|^2 + \alpha tr(\mathbf{V}^T \mathbf{L}^F \mathbf{V}) + \beta tr(\mathbf{U}^T \mathbf{L}^L \mathbf{U}) \tag{1}$$

$\mathbf{U} \in \mathbb{R}^{q \times r} \geq 0$ and $\mathbf{V} \in \mathbb{R}^{n \times r} \geq 0$ are the low-dimensional representations of the original $\mathbf{Y}$ matrix based on rows and columns, respectively ($r \ll q$, $r \ll n$). α and β are the positive scalar parameters. By integrating two regularizations into NMF, MF-INL enforces the factorized low-rank matrices to preserve the geometric structure of instances and the interrelationships of labels in a coherent and coordinated manner. We will elaborate on the two regularization constraints in the following subsections.

3.1 Feature-based regularization

Features of an instance essentially decide its outputs (labels) [11,30]. To leverage the geometric structure of instances, which depends on the instance features, we adopt the manifold assumption, which is widely-used in dimensionality reduction and semi-supervised learning [22,1]. The manifold assumption assumes that if two data points $\mathbf{x}_i$ and $\mathbf{x}_j$ are close in the intrinsic geometry of the ambient space, they should have similar outputs (or labels). Firstly, we construct a k nearest neighbor graph [10,9] to model the local geometric structure of n instances as follows:

$$\mathbf{W}_{ij}^F = \begin{cases} 1, \text{if } \mathbf{x}_j \in \mathcal{N}_k(\mathbf{x}_i) \\ 0, \text{ otherwise} \end{cases} \tag{2}$$

where $\mathcal{N}_k(\mathbf{x}_i)$ is the set of k nearest neighbors of $\mathbf{x}_i$, and the neighborhood relationship is determined using the Euclidean distance. Here, the 0-1 weighting scheme is used to weighting the edges of the kNN graph, but other weighting schemes and distance metrics can also be defined based on the specific application domains. $\mathbf{v}_i$ is the low-dimensional representation of i-th column of $\mathbf{Y}$. The manifold assumption to enforce that $\mathbf{v}_i$ and

$\mathbf{v}_j$ are close to each other when $\mathbf{x}_i$ and $\mathbf{x}_j$ are close in the original ambient space is specified as follows:

$$\psi_1(\mathbf{V}) = \frac{1}{2}\sum_{i,j=1}^{n} \|\mathbf{v}_i - \mathbf{v}_j\|^2 \mathbf{W}_{ij}^F = tr(\mathbf{V}^T\mathbf{L}^F\mathbf{V}) \tag{3}$$

where $\mathbf{D}^F$ is a diagonal matrix with $\mathbf{D}_{ii}^F = \sum_{j=1}^{n} \mathbf{W}_{ij}^F$, $\mathbf{L}^F = (\mathbf{D}^F - \mathbf{W}^F)$ is the graph Laplacian matrix, and $tr(\cdot)$ denotes the trace of a matrix. By minimizing $\psi_1(\mathbf{V})$, the proximity between close instances in the original space can be preserved in the low-dimensional space spanned by $\mathbf{V}$ through the feature-based regularization.

3.2 Label-based regularization

In multi-label learning, labels are not mutually exclusive, and different pairs of labels (or groups of labels) may have different degrees of correlation. In contrast, traditional multi-class classification and partial-label learning explicitly (or implicitly) assumes the labels are mutually exclusive. Various types of label correlations have been explored in multi-label learning and they generally can improve the performance [33]. Given this, besides feature-aware constraint, we also define a label-aware constraint.

Each row of $\mathbf{U}$ can be viewed as a low-dimensional (or latent) label vector of the original label vector expressed by the corresponding row of $\mathbf{Y}$. The expectation is that the new low-dimensional representation is able to preserve the interrelationship of labels in the original space, that is, if labels s and t are correlated in $\mathbf{Y}$, then the low-dimensional representations of $\mathbf{Y}_s$ and $\mathbf{Y}_t$, i.e. $\mathbf{u}_s$ and $\mathbf{u}_t$, would also be correlated in the latent label space. We first measure the label correlation using the cosine similarity as follows:

$$\mathbf{W}_{st}^L = \frac{\mathbf{Y}_s\mathbf{Y}_t^T}{\|\mathbf{Y}_s\|\|\mathbf{Y}_t\|} \tag{4}$$

where $\mathbf{Y}_s$ is the s-th row of $\mathbf{Y}$, and $\mathbf{W}_{st}^L \in [0,1]$ denotes the label correlation between the s-th and t-th labels. $\mathbf{W}_{st}^L$ is large when s and t frequently co-occur as annotation of instances, and is small otherwise. We use cosine similarity for its simplicity, but other measures can be also used. To preserve label correlations in the latent label space, we defines the label-aware constraint for $\mathbf{U}$ as follows:

$$\psi_2(\mathbf{U}) = \frac{1}{2}\sum_{s,t=1}^{q} \|\mathbf{u}_s - \mathbf{u}_t\|^2 \mathbf{W}_{st}^L = tr(\mathbf{U}^T\mathbf{L}^L\mathbf{U}) \tag{5}$$

where $\mathbf{D}^L$ is a diagonal matrix with $\mathbf{D}_{st}^L = \sum_{t=1}^{q} \mathbf{W}_{st}^L$ and $\mathbf{L}^L = \mathbf{D}^L - \mathbf{W}^L$. As in Eq.(3), by minimizing Eq.(5), we can preserve correlations in the latent label space.

As for the standard NMF, we optimize Eq.(1) using an iterative algorithm. Readers can refer to [2] for details. By minimizing Eq.(1), MF-INL reconstructs the approximated association matrix as $\hat{\mathbf{Y}} = \mathbf{U}\mathbf{V}^T$. After the reconstruction, the associations available in $\mathbf{Y}$, but inconsistent with the low-rank representation with respect to the geometric structure of instances and to the correlations between labels, have low values in $\hat{\mathbf{Y}}$; otherwise have high values. As a result, each entry of $\hat{\mathbf{Y}}$ reflects the association confidence between a particular instance and a particular label. The labels corresponding to the smaller entries in $\hat{\mathbf{Y}}$ are more likely to be deemed as noisy labels.

4 Experimental setup

Datasets: To study the performance of MF-INL, we conduct experiments on four multi-label datasets (listed in Table 1, downloaded from the Mulan Library[3].). Since there are no off-the-shelf multi-label datasets that can be directly used to validate the performance of identifying noisy labels of multi-label instances, we assume that the available labels of instances in these four datasets are noise-free, and randomly inject additional $p \times q$ labels to each instance as noisy labels, where p is the ratio of noisy labels. Specially, to study the performance of MF-INL under different levels of noise, we conduct experiments with p set to 0.3 and 0.5.

Table 1: Datasets used in the experiments. Avg-Labels is the average number of labels per instance.

Dataset	Instances	Features	Labels	Avg-Labels
Enron	1702	1001	53	3.378
Yeast	2417	103	14	4.237
Rcv1-s5	6000	47235	101	2.642
Tmc	28596	500	22	2.220

Comparative methods: We compare MF-INL against IPAL [31], PL-LEAF [34], CPLST [3], FaIE[15], ProDM [27]. These methods have been presented in the Section 2.

Evaluation metrics: We use three representative multi-label learning and partial-label learning evaluation metrics: RankingLoss (RL), OneError (OE), and AveragePrecision (AP) [36]. They can evaluate the identification of noisy labels from the perspective of label distribution [5]. Note that the smaller the values of RL and OE, the better the performance is; while the larger the values of AP, the better the performance is. We report the *1-RL* and the *1-OE* in the following experiments. As such, *larger* values imply a *better* performance.

5 Experimental results and analysis

5.1 Noisy label identification

Table 2: Performance for the identification of noisy labels as the ratio (p) of randomly injected noisy labels per instance increases. •/○ indicates whether MF-INL is statistically (according to a pairwise t-test at 95% significance level) superior/inferior to the other method for a particular value of p.

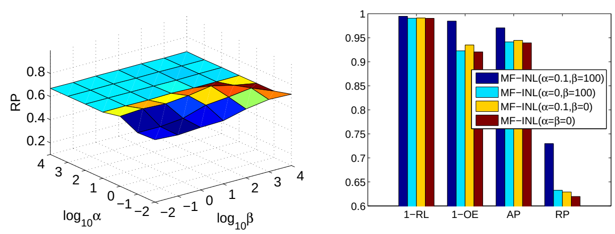
		p	CPLST	FaIE	ProDM	IPAL	PL-LEAF	MF-INL
Enron	1-RL	0.3	0.799 ± 0.006•	0.776 ± 0.003•	0.996 ± 0.000•	0.992 ± 0.000•	0.994 ± 0.000•	0.998 ± 0.000
		0.5	0.812 ± 0.007•	0.784 ± 0.004•	0.995 ± 0.000•	0.987 ± 0.000•	0.990 ± 0.000•	0.996 ± 0.000
	1-OE	0.3	0.973 ± 0.004•	0.971 ± 0.003•	0.964 ± 0.003•	0.922 ± 0.004•	0.982 ± 0.003•	0.985 ± 0.004
		0.5	0.963 ± 0.004•	0.960 ± 0.003•	0.936 ± 0.004•	0.868 ± 0.005•	0.966 ± 0.000•	0.970 ± 0.005
	AP	0.3	0.801 ± 0.006•	0.780 ± 0.003•	0.951 ± 0.002•	0.910 ± 0.002•	0.943 ± 0.000•	0.973 ± 0.002
		0.5	0.804 ± 0.007•	0.779 ± 0.004•	0.930 ± 0.002•	0.853 ± 0.002•	0.909 ± 0.001•	0.953 ± 0.002
Yeast	1-RL	0.3	0.970 ± 0.004•	0.975 ± 0.001•	0.977 ± 0.001•	0.960 ± 0.001•	0.966 ± 0.000•	0.985 ± 0.001
		0.5	0.936 ± 0.006•	0.932 ± 0.003•	0.956 ± 0.001•	0.930 ± 0.001•	0.943 ± 0.001•	0.970 ± 0.001
	1-OE	0.3	0.971 ± 0.006○	0.976 ± 0.003○	0.941 ± 0.005•	0.888 ± 0.004•	0.917 ± 0.002•	0.967 ± 0.003
		0.5	0.885 ± 0.016•	0.885 ± 0.011•	0.914 ± 0.004•	0.831 ± 0.003•	0.893 ± 0.001•	0.945 ± 0.004
	AP	0.3	0.955 ± 0.005•	0.960 ± 0.003•	0.952 ± 0.002•	0.914 ± 0.002•	0.930 ± 0.001•	0.969 ± 0.001
		0.5	0.892 ± 0.008•	0.884 ± 0.008•	0.919 ± 0.002•	0.867 ± 0.002•	0.896 ± 0.001•	0.946 ± 0.003
Tmc	1-RL	0.3	0.903 ± 0.004•	0.879 ± 0.001•	0.995 ± 0.000	0.989 ± 0.000•	0.992 ± 0.000•	0.995 ± 0.000
		0.5	0.903 ± 0.002•	0.879 ± 0.002•	0.991 ± 0.000○	0.980 ± 0.000•	0.991 ± 0.000	0.990 ± 0.001
	1-OE	0.3	0.929 ± 0.005•	0.937 ± 0.002•	0.968 ± 0.001•	0.869 ± 0.001•	0.993 ± 0.000○	0.987 ± 0.001
		0.5	0.885 ± 0.003•	0.894 ± 0.005•	0.917 ± 0.001•	0.765 ± 0.001•	0.979 ± 0.000○	0.939 ± 0.005
	AP	0.3	0.886 ± 0.004•	0.872 ± 0.001•	0.971 ± 0.000	0.929 ± 0.000•	0.955 ± 0.000•	0.971 ± 0.002
		0.5	0.860 ± 0.002•	0.849 ± 0.003•	0.940 ± 0.001○	0.868 ± 0.001•	0.959 ± 0.000○	0.937 ± 0.004
Rcv1-s5	1-RL	0.3	0.584 ± 0.003•	0.573 ± 0.002•	0.998 ± 0.000•	0.998 ± 0.000•	0.996 ± 0.000•	0.999 ± 0.000
		0.5	0.591 ± 0.004•	0.577 ± 0.002•	0.997 ± 0.000•	0.996 ± 0.000•	0.993 ± 0.000•	0.998 ± 0.000
	1-OE	0.3	0.881 ± 0.001•	0.881 ± 0.001•	0.923 ± 0.004•	0.943 ± 0.002•	0.907 ± 0.000•	0.992 ± 0.002
		0.5	0.878 ± 0.001•	0.876 ± 0.001•	0.875 ± 0.004•	0.886 ± 0.003•	0.868 ± 0.000•	0.978 ± 0.004
	AP	0.3	0.589 ± 0.003•	0.578 ± 0.002•	0.948 ± 0.002•	0.945 ± 0.001•	0.911 ± 0.000•	0.979 ± 0.001
		0.5	0.592 ± 0.004•	0.577 ± 0.002•	0.915 ± 0.002•	0.902 ± 0.002•	0.872 ± 0.000•	0.963 ± 0.002

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

Following the experimental protocol in partial-label learning [34], we considered all instances in each dataset as both training and testing data. We search the optimal parameter values for α and β in the range $\{0, 0.01, 0.1, 1, 10, 100, 1000, 10000\}$. As a result, α and β are set to 0.1 and 100, respectively. The neighborhood size k is set to 5. The same value $r = 10$ is used for MF-INL, CPLST, and FaIE. Other parameters of comparing methods are set (or optimized) as suggested by the authors in their code, or respective papers. Table 2 reports the average results of 10 independent runs of all methods under each particular p. From Table 2, MF-INL outperforms the other methods across all the metrics in most cases. This observation shows that MF-INL can identify noisy labels of multi-label instances more accurately than other related methods, and supports our motivation to factorize the instance-label association matrix into latent low-rank subspaces.

To study the efficiency of MF-INL, we record the runtime of all other comparing methods on a server with configuration: CentOS 7, 256GB RAM and Intel Exon E5-2678v3. The total runtime (seconds) of CPLST, FaIE, ProDM, IPAL, PL-LEAF and MF-INL on all datasets is 173, 208, 306, 2933, 1962475 and 108, respectively. From these observations and the results in Table 2, we can conclude that MF-INL not only holds comparable runtime against efficient counterparts, but also achieves a superior performance.

5.2 Parameter sensitivity analysis



(a) Tmc (b) Tmc

Fig. 1: *RP* of MF-INL under different combinations of α and β on Enron and Tmc.

To investigate the sensitivity of α and β, we vary α and β in the range $\{0.01, 0.1, 1, 10, 100, 1000, 10000\}$ with $p = 0.3$, and report the average RP of MF-INL in 10 independent runs under different combinations of α and β in Fig 1(a). From this figure, we can see that MF-INL achieves a stable and good performance for a wide range of α and β values. In addition, we can see that MF-INL, with values $\alpha = 0.01$ and $\beta = 0.01$, has lower RP than many other values' combinations. This observation suggests that it's necessary to integrate feature-based and label-based regularizations into NMF to obtain coherent matrices. To further investigate the influence of feature-based and label-based regularizations, we test the performance of MF-INL under extreme settings of α and β, that is: $\alpha = 0$ and $\beta = 100$; $\alpha = 0.1$ and $\beta = 0$; and We report the results of MF-INL under these extreme settings in Fig. 1(b). From the results we can see that using feature-based and label-based regularizations together can significantly improve the performance of NMF. Using either one of the two regularizations gives comparable or better performance than NMF alone. $\alpha = 0$ and $\beta = 0$.

The rank size of the decomposed (projected) matrix is an essential parameter for MF-INL and LSDR-based methods CPLST and FaIE. We also conduct experiments to study the sensitivity of r. Due to page limitation, we do not provide the results in this paper. From the results, MF-INL is robust to different input values of r, while CPLST and FaIE are sensitive to r. Besides, MF-INL outperforms CPLST and FaIE under each considered value of r. The robustness of MF-INL to r can be attributed to the fact that MF-INL can find coherent low-rank matrices by simultaneously preserving the geometric structure of instances and the correlation of labels.

6 Conclusions and future work

In this paper, we study an interesting but rarely explored problem of multi-label learning: identifying noisy labels of multi-label instances. To solve this problem, we introduce a matrix factorization based method called MF-INL. The experimental study shows that MF-INL can identify noisy labels more accurately than other competitive techniques. It will be interesting to study the performance of MF-INL under different choices of distance metrics and label correlations, and to iteratively update the correlations, since label correlations are affected by noise in the original instance-label association matrix.

Acknowledgments

This work is supported by Natural Science Foundation of China (61751217 and 61402378), Natural Science Foundation of CQ CSTC(cstc2016jcyjA0351), Open Research Project of Hubei Key Laboratory of Intelligent Geo-Information Processing (KLIGIP-2017A05) and Chongqing Graduate Student Research Innovation Project [No. CYS18070].

References

1. Belkin, M., Niyogi, P., Sindhwani, V.: Manifold regularization: A geometric framework for learning from labeled and unlabeled examples. JMLR **7**(11), 2399–2434 (2006)
2. Cai, D., He, X., Han, J., Huang, T.S.: Graph regularized nonnegative matrix factorization for data representation. TPAMI **33**(8), 1548–1560 (2011)
3. Chen, Y., Lin, H.: Feature-aware label space dimension reduction for multi-label classification. In: NIPS. pp. 1529–1537 (2012)
4. Cour, T., Sapp, B., Taskar, B.: Learning from partial labels. JMLR **12**(5), 1501–1536 (2011)
5. Geng, X.: Label distribution learning. TKDE **28**(7), 1734–1748 (2016)
6. Gibaja, E., Ventura, S.: A tutorial on multilabel learning. ACM Computing Surveys **47**(3), 52 (2015)
7. Hansen, P.C., Jensen, S.H.: Fir filter representations of reduced-rank noise reduction. IEEE Transactions on Signal Processing **46**(6), 1737–1741 (1998)
8. Hüllermeier, E., Beringer, J.: Learning from ambiguously labeled examples. Intelligent Data Analysis **10**(5), 419–439 (2006)
9. Jiang, L., Wang, D., Cai, Z., Jiang, S., Yan, X.: Scaling up the accuracy of k-nearest-neighbour classifiers: A naïve-bayes hybrid. International Journal of Computers and Applications **31**(1), 36–43 (2009)
10. Jiang, L., Cai, Z., Wang, D., Zhang, H.: Bayesian citation-knn with distance weighting. International Journal of Machine Learning and Cybernetics **5**(2), 193–199 (2014)
11. Jiang, L., Zhang, L., Li, C., Wu, J.: A correlation-based feature weighting filter for naive bayes. TKDE, DOI: 10.1109/TKDE.2018.2836440 (2018)
12. Konstantinides, K., Natarajan, B., Yovanof, G.S.: Noise estimation and filtering using block-based singular value decomposition. IEEE Transactions on Image Processing **6**(3), 479–483 (1997)

13. Lee, D.D., Seung, H.S.: Algorithms for non-negative matrix factorization. In: NIPS. pp. 556–562 (2001)
14. Li, C., Sheng, V.S., Jiang, L., Li, H.: Noise filtering to improve data and model quality for crowdsourcing. Knowledge-Based Systems **107**, 96–103 (2016)
15. Lin, Z., Ding, G., Hu, M., Wang, J.: Multi-label classification via feature-aware implicit label space encoding. In: ICML. pp. 325–333 (2014)
16. Liu, L., Dietterich, T.G.: A conditional multinomial mixture model for superset label learning. In: NIPS. pp. 548–556 (2012)
17. Meng, D., De La Torre, F.: Robust matrix factorization with unknown noise. In: ICCV. pp. 1337–1344 (2013)
18. Nam, J., Kim, J., Mencía, E.L., Gurevych, I., Fürnkranz, J.: Large-scale multi-label text classificationłrevisiting neural networks. In: ECML. pp. 437–452 (2014)
19. Sun, Y., Zhang, Y., Zhou, Z.: Multi-label learning with weak label. In: AAAI. pp. 593–598 (2010)
20. Tai, F., Lin, H.: Multilabel classification with principal label space transformation. Neural Computation **24**(9), 2508–2542 (2012)
21. Tang, C., Zhang, M.: Confidence-rated discriminative partial label learning. In: AAAI. pp. 2611–2617 (2017)
22. Van Der Maaten, L., Postma, E., Van den Herik, J.: Dimensionality reduction: a comparative review. JMLR **10**, 66–71 (2009)
23. Wu, B., Lyu, S., Hu, B.G., Ji, Q.: Multi-label learning with missing labels for image annotation and facial action unit recognition. Pattern Recognition **48**(7), 2279–2289 (2015)
24. Xu, C., Tao, D., Xu, C.: Robust extreme multi-label learning. In: KDD. pp. 1275–1284 (2016)
25. Yeh, C., Wu, W., Ko, W., Wang, Y.F.: Learning deep latent space for multi-label classification. In: AAAI. pp. 2838–2844 (2017)
26. Yu, F., Zhang, M.L.: Maximum margin partial label learning. Machine Learning **104**(4), 573–593 (2017)
27. Yu, G., Domeniconi, C., Rangwala, H., Zhang, G.: Protein function prediction using dependence maximization. In: ECML/PKDD. pp. 574–589 (2013)
28. Yu, G., Zhang, G., Rangwala, H., Domeniconi, C., Yu, Z.: Protein function prediction using weak-label learning. In: ACM Conference on Bioinformatics, Computational Biology and Biomedicine. pp. 202–209 (2012)
29. Zhang, J., Wu, X., Sheng, V.S.: Learning from crowdsourced labeled data: a survey. Artificial Intelligence Review **46**(4), 543–576 (2016)
30. Zhang, L., Jiang, L., Li, C.: A new feature selection approach to naive bayes text classifiers. International Journal of Pattern Recognition and Artificial Intelligence **30**(02), 1650003 (2016)
31. Zhang, M., Yu, F.: Solving the partial label learning problem: an instance-based approach. In: IJCAI. pp. 4048–4054 (2015)
32. Zhang, M., Yu, F., Tang, C.: Disambiguation-free partial label learning. TKDE **29**(10), 2155–2167 (2017)
33. Zhang, M., Zhang, K.: Multi-label learning by exploiting label dependency. In: KDD. pp. 999–1008 (2010)
34. Zhang, M., Zhou, B., Liu, X.: Partial label learning via feature-aware disambiguation. In: KDD. pp. 1335–1344 (2016)
35. Zhang, M., Zhou, Z.: Ml-knn: A lazy learning approach to multi-label learning. Pattern Recognition **40**(7), 2038–2048 (2007)
36. Zhang, M., Zhou, Z.: A review on multi-label learning algorithms. TKDE **26**(8), 1819–1837 (2014)