Partial Multi-Label Learning with Global and Local Manifold Disambiguation

Qizheng Pan[1*], Jianmin Li[2], Ying Ma[3]

[1,2,3]*College of Computer and Information Engineering, Xiamen University of Technology, Jimei District, Xiamen, China*
panqizheng@s.xmut.edu.cn, {lijm, maying}@xmut.edu.cn
**Corresponding author. Email: panqizheng@s.xmut.edu.cn*

Abstract
In Partial Multi-Label learning (PML), each training example is assigned with a candidate label set where only partial labels are correct. Existing PML methods only focus on global label correlation, while they lack the consideration of the local label correlation. To alleviate this issue, a novel framework is proposed to jointly consider the feature manifold structure over the global instances and local instances. Specifically, we firstly explore the global feature manifold and local feature manifold by the affinity information conveyed by feature vectors. Then, a trade-off parameter is introduced to character the relative contribution of the feature manifold structures optimized by different methods. Afterwards, in order to disambiguate the candidate labels, we utilize the joint feature manifold in the label space. Finally, the predicted results are learned by training a linear multi-label classification model. Extensive experiments on six PML datasets demonstrate the effectiveness of our proposed method.

Keywords: *Partial Multi-Label Learning, Disambiguation, Manifold*

1 INTRODUCTION

Unlike traditional supervised learning framework, Partial Multi-Label Learning (PML) is a weakly supervised learning framework, in which each example is associated with candidate labels but only partial of which are correct. Since human annotators have difficult to giving the accurate labels when the human annotators are careless or weary [13], some images cannot be classified accurately. For example, as shown in Figure 1, the set of candidate labels are given by human annotators. Among the eight candidate labels, only some of them (Mountain, House, Car, Cloud, Tree) are correct, whereas other labels (Boat, People, Animal) are incorrect due to the potential untrustworthy annotators. The task of PML methods is to train a multi-label predictor from PML training data with uncertain labels and accurately predict the unseen examples.

The main challenge for PML is how to disambiguate from the candidate labels and discriminate the ground-truth labels. To tackle this problem, the PML methods can be divided into two strategies. For the first strategy, the whole process is separated into two-stage. For example, [9] proposed an effective approach PARTICES, where the credible labels with high confidences are recognized by iterative label propagation at the first stage and then a multi-label predictor is learned via virtual label splitting (PARTICE-VLS) or MAP reasoning (PARTICE-MAP) at the second stage. DRAMA [7] is also a novel two-stage method, where a label confidence matrix is generated for disambiguation by utilizing the feature manifold, followed by a gradient boosting-based model that influences the feature space in each boosting round. For the second strategy, the whole process is unified. For example, PML-FP is proposed by [8], where a classification predictor along with label confidences are optimized in a coherent framework by minimizing the confidence weighted ranking loss and exploiting data information. One recent work [6] is another popular PML approach, which incorporates example-label assignment correlations, as well as example correlations and label correlations, into the proposed framework at the same time.

B. Fox et al. (Eds.): IC-ICAIE 2022, AHCS 9, pp. 1364-1370, 2023.
https://doi.org/10.2991/978-94-6463-040-4_203

Figure 1: The image is partially tagging by human annotators. Among the eight candidate labels, only five of them in black font are ground-truth labels including mountain, house, car, cloud and tree, whereas three of them are false including boat, animal and people.

We find that current PML methods only consider either global feature manifold or label correlations alone, whereas the global feature manifold and local feature manifolds are hardly taken into consideration together to induce the label manifold. However, the local feature manifold can better help to explore the local label correction. For example, instances containing both "Buck" and "Tiger" labels are often appear in a local data subset with the semantic "Animal", while instances containing both "Buck" and "Laker" often appear in another data subset with the semantic "Basketball Teams". Meanwhile, according to the smoothness assumption [1], closely related instances are more inclined to share same labels, implying that the local topological structure can be transferred from the feature space to the label space [4]. Based on this assumption, the local label subsets with the similar semantic have more similar labels. According to experimental results, considering global and local feature manifolds simultaneously is more beneficial.

In this paper, we introduce a novel approach called Partial Multi-Label Learning with Global and LOcal MANifold disambiguation (GLOMAN). To better explored the relation between labels, we consider not only the global feature manifold but also the local feature manifold. Specifically, motivated by [5], we firstly separate our training data into multiple groups, with each group sharing a portion of label correlations. The groups can be generated by clustering algorithm. Then, we explore the local feature manifold in a group by an affinity graph reconstructed by each instance and its K-nearest neighbors. Afterwards, the local affinity information from different groups is fused into an aggregate feature manifold structure. After that, the global affinity information is directly explored over all instances. Secondly, we simply use a trade-off parameter to characterize the relative contribution of the both affinity information achieved by different ways and gain a joint affinity graph. Then, based on the smoothness assumption, we utilize the joint affinity graph to induce the manifold in the label space. Finally, a linear multi-label classification model is learned to predict the unseen instances.

2 PROPOSED METHOD

Specifically speaking, $X = \mathbb{R}^{d\times m}$ represents the d dimensional feature space for m instances and $\mathcal{Y} = \{y_1, y_2, \dots, y_q\}$ with q class labels denotes the label space. The goal of PML methods aims to train a multi-label classifier $f: X \mapsto 2^{\mathcal{Y}}$ based on the PML training data $\mathcal{D} = \{(\boldsymbol{x}_i, Y_i) \mid 1 \le i \le m\}$, in which the instance $\boldsymbol{x}_i$ is described a d-dimensional feature vectors, and $Y_i \subseteq \mathcal{Y}$ denotes the candidate labels associated with the instance $\boldsymbol{x}_i$.

Firstly, GLOMAN aims to explore the feature manifold structure over all training instances. Given an affinity Graph $G = (V, E, \mathbf{S_G})$ based on the PML dataset $\mathcal{D}$, $V = \{\boldsymbol{x}_i \mid 1 \le i \le m\}$ is equivalent to the set of training instances and $E = \{(\boldsymbol{x}_i, \boldsymbol{x}_j) \mid i \in kNN(\boldsymbol{x}_j), 1 \le j \le m\}$ is equivalent to the edge set in which $kNN(\boldsymbol{x}_j)$ denotes the index set of $\boldsymbol{x}_j$'s k-nearest neighbours. $\boldsymbol{S_G} = [s_{ij}]_{m\times m}$ is a affinity matrix which preserves the affinity information conveyed by feature vectors. For each instance $\boldsymbol{x}_i$, it can be linearly reconstructed from its nearest neighbours and the value s_{ij} of $\boldsymbol{S_G}$ is optimized by sovling follow minimum error reconstruction problems:

$$\begin{aligned} &\min_{\mathbf{S_G}} \sum_{j=1}^{m} \|x_j - \sum_{(x_i,x_j)\in E} s_{ij}\cdot x_i\|_2^2 \\ &\text{s.t.:} \sum_{(x_i,x_j)\in E} s_{ij} = 1 (1 \le j \le m) \\ &\quad s_{ij} \ge 0 (\forall (x_i, x_j) \in E) \\ &\quad s_{ij} = 0 (\forall (x_i, x_j) \notin E) \end{aligned} \tag{1}$$

Conceptually, the resulting of (1) can be seen as a standard quadratic programming QP problem, which can be solved by any off-the-shelf quadratic programming QP solvers. By solving (1), the global feature manifold can be effectively exploited.

Since local label correlations can only be affected by a local group, we consider the local feature manifold accordingly. As depicted in Figure 2, we assume that the dataset X can be separated into g groups $\{\boldsymbol{X}_1, \dots, \boldsymbol{X}_g\}$ via a clustering algorithm such as k-means. It is noteworthy that each instance in the group can only be linearly reconstructed by its k-nearest neighbors in this group and then the s_{ij} optimized in every group are also aggregated into a $m \times m$ affinity matrix. Compared to the S_G, the affinity matrix S_L is obtained by the local data subset instead of the global dataset.

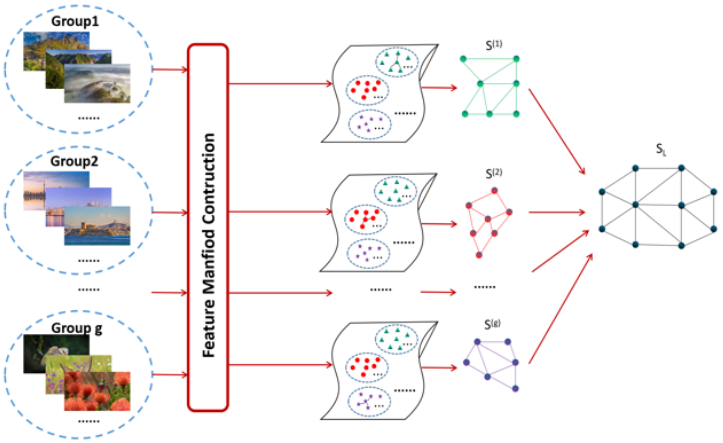

Figure 2: The aggregate feature manifold is fused by every local feature manifold. For convenience, we simply describe the affinity graph of each group.

By considering the local feature manifold, let $E^u = \{(\boldsymbol{x}_i, \boldsymbol{x}_j) \mid i \in kNN(\boldsymbol{x}_j), 1 \le j \le m^u\}$ be the sub edge set of E and m^u be to the number of instances in the group u. Then we have the following optimization problem:

$$\begin{aligned} &\min_{\mathbf{S_L}} \sum_{u=1}^{g} \sum_{j=1}^{m_u} \|\boldsymbol{x}_j - \sum_{(x_i,x_j)\in E^u} s_{ij} \cdot \boldsymbol{x}_i\|_2^2 \\ &\text{s.t.:} \sum_{u=1}^{g} \sum_{(x_i,x_j)\in E^u} s_{ij} = 1(1 \le j \le m_u) \\ &\qquad s_{ij} \ge 0(\forall (x_i, x_j) \in E) \\ &\qquad s_{ij} = 0(\forall (x_i, x_j) \notin E) \end{aligned} \tag{2}$$

Similar to (1), the (2) is also a standard QP problem which can also be solved by applying off-the-shelf QP toolbox. Global feature manifold and local feature manifold are encoded in the affinity matrix S_G *and* S_L. By taking the basic assumption that global feature manifold and local feature manifold have different contribution for model inducing, we use a trade-off parameter α $(0 \le \alpha \le 1)$ to characterize the relative contribution to the joint feature manifold structure:

$$S = (1-\alpha)S_G + \alpha S_L (0 \le \alpha \le 1) \tag{3}$$

Where S denote the joint feature manifold by fusing S_G *and* S_L.

According to smoothness assumption, similar examples in the feature space are more inclined to share the same topology in the label space. GLOMAN aims to disambiguate the irrelevant labels in a numerical label space, among which the values of the numerical label can be regarded as the labeling confidence. Formally, the label manifold is reconstructed by the following minimization:

$$\begin{aligned} &\min_{\mathrm{F}} \sum_{j=1}^{m} \|f_j - \sum_{i=1}^{m} s_{ij} \cdot f_i\|_2^2 \\ &\text{s.t.} : \sum_{l=1}^{q} f_{jl} = 1(1 \le j \le m) \\ &\qquad f_{jl} \ge 0(1 \le j \le m, y_l \in Y_j) \\ &\qquad f_{jl} = 0(1 \le j \le m, y_l \notin Y_j) \end{aligned} \tag{4}$$

Here, $\mathbf{F} = [f_1, f_2, \ldots, f_m]^\top = \{f_{jl}\}$ denotes the labeling confidence matrix, where f_{jl} denotes the confidence score of label y_l being the ground-truth label of x_j.

To achieve the better solution of confidence scores matrix F with the off-the-shelf QP toolbox, we convert the (4) into (5).

$$\begin{aligned} &\min_{\tilde{f}} \frac{1}{2} \tilde{f}^\top \Psi \tilde{f} \\ &\text{s.t. } \mathbf{0}_{m\cdot q} \le \tilde{f} \le \overline{\boldsymbol{y}}, E\tilde{f} = \mathbf{1}_m \end{aligned} \tag{5}$$

Here, $\tilde{f} \in \{0,1\}^{m\cdot q}$ is the vector-wise of F and $\overline{\boldsymbol{y}} \in \{0,1\}^{m\cdot q}$ is the vector-wise of the candidate labels Y. Moreover, $\mathbf{E} = [\mathbf{I}_{m\times m}, \mathbf{I}_{m\times m}, \ldots, \mathbf{I}_{m\times m}] \in \{0,1\}^{m\times m\cdot q}$ and $\mathbf{1}_m \in R^{m\times 1}$. Then, we define $\mathrm{T} = \mathrm{S}^\top \mathrm{S} + (\mathrm{S}^\top \mathbf{1}_{m\times m} \mathrm{S}) \odot \mathrm{I}_{m\times m} - 2\mathrm{S}$, where $\odot$ is the the Hadamard product and $\mathrm{I}_{m\times m}$ is an identity matrix. In addition, the Ψ is calculated by $\psi + \psi^\top$, where $\psi \in R^{m\cdot q\times m\cdot q}$ is defined as:

$$\psi = \begin{bmatrix} \mathrm{T} & 0_{m\times m} & \cdots & 0_{m\times m} \\ 0_{m\times m} & \mathrm{T} & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0_{m\times m} \\ 0_{m\times m} & \cdots & 0_{m\times m} & \mathrm{T} \end{bmatrix} \tag{6}$$

The (5) is a standard QP problem, which can also be solved by any off-the-shelf QP toolbox. By solving (5),

the confidence value can be achieved to disambiguate the noise labels.

Let $D = \{(\boldsymbol{x}_i, \hat{Y}_i) \mid 1 \le i \le m\}$ be the disambiguated counterpart, the ground-truth labels $\hat{Y}$ reformed by a threshold parameter θ_1:

$$\hat{Y}_i = \{y_l \mid f_{jl} \ge \theta_1, 1 \le l \le q\} \tag{7}$$

The corresponding $y_l = 1$ when the confidence value $f_{il} \ge \theta_1$, $y_l = 0$.

Finally, GLOMAN intends to train a linear multi-label predictive model to make prediction on unseen instances. By solving the following ridge regression problem, the predictive model can be learned:

$$\min_W \frac{1}{2}\|\hat{Y} - W^{\top}\mathrm{X}\|_F^2 + \gamma \cdot \|W\|_F^2 \tag{8}$$

Here, $\mathrm{X} \in R^{d\times m}$ and $\hat{\mathrm{Y}} \in R^{q\times m}$ represents the instances matrix and the ground-truth labels matrix respectively. Accordingly, $W = [w_1, w_2, \dots, w_q] \in R^{d\times q}$ denotes the linear predictive matrix and γ denotes the regularization coefficient. To control model complexity, we apply the commonly utilized squared Frobenius norm for regularization. The solution of ridge regression can be solved with standard equation method:

$$W = (\mathrm{X}\mathrm{X}^{\top} + \gamma \cdot \mathrm{I})^{-1}\mathrm{X}\hat{Y}^{\top} \tag{9}$$

Therefore, for unseen instance $\tilde{x} \in \mathrm{X}$, a threshold parameter θ_2 is introduced to predict the modeling outputs $\tilde{Y}$:

$$\tilde{Y} = \{y_j \mid w_j^{\top} \cdot \tilde{x} \ge \theta_2, 1 \le j \le q\} \tag{10}$$

3 EXPERIMENTS

3.1 Experimental Setup

We conduct experiments on widely-used 6 synthetic PML datasets to evaluate the performance of our proposed GLOMAN approaches. Following the [12], we use the same protocol for adding label noise where the PML candidate label sets are produced from the original MLL datasets by

Table 1: Characteristics of the partial multi-label datasets.

Dataset	N(E)	N(F)	N(L)	Avg(L)	Domains
Emotions	593	72	6	1.87	music
Medical	978	1449	45	1.25	text
Image	2000	294	6	1.23	images
Scene	2407	294	6	1.07	images
Yeast	2417	103	14	4.23	biology
Bibtex	7395	1836	159	2.40	text

Table 2: Comparison of GLOMAN with well-established MLL and PML approaches on five evaluation measurements, where the best performances are shown in bold face (p=7, r=3).

Datasets	ML-KNN	RANK-SVM	PARTICE-VLS	PARTICE-MAP	DRAMA	GLOMAN
Hamming Loss ↓						
Emotions	0.635±0.025	0.605±0.018	0.254±0.020	0.294±0.033	0.364±0.011	**0.240±0.027**
Medical	0.016±0.002	0.038±0.002	0.026±0.003	0.037±0.004	0.013±0.000	**0.012±0.002**
Image	0.371±0.040	0.409±0.018	0.217±0.056	0.258±0.094	0.261±0.003	**0.197±0.011**
Scene	0.252±0.017	0.460±0.043	0.141±0.032	0.175±0.047	0.249±0.002	**0.116±0.011**
Yeast	0.209±0.011	0.246±0.010	**0.207±0.005**	0.233±0.005	0.231±0.001	0.245±0.008
Bibtex	0.014±0.000	-	0.015±0.000	0.024±0.001	**0.012±0.000**	0.015±0.000
Ranking Loss ↓						
Emotions	0.366±0.031	0.415±0.045	0.266±0.020	0.268±0.039	0.330±0.008	**0.199±0.040**
Medical	0.068±0.018	0.139±0.016	0.106±0.030	0.108±0.024	0.038±0.000	**0.032±0.011**
Image	0.240±0.017	**0.201±0.017**	0.239±0.061	0.296±0.094	0.215±0.004	0.207±0.014
Scene	0.144±0.012	0.387±0.021	0.130±0.046	0.248±0.058	**0.111±0.001**	**0.111±0.014**
Yeast	0.177±0.012	**0.166±0.012**	0.185±0.011	0.181±0.012	0.223±0.000	0.191±0.013
Bibtex	0.230±0.010	-	0.310±0.009	0.329±0.008	**0.139±0.000**	0.147±0.006
One Error ↓						
Emotions	0.518±0.061	0.641±0.094	0.329±0.039	0.410±0.063	0.422±0.020	**0.320±0.060**
Medical	0.272±0.032	0.724±0.040	0.244±0.062	0.517±0.075	0.175±0.005	**0.155±0.040**

Image	0.404±0.027	**0.340±0.031**	0.360±0.106	0.454±0.152	0.341±0.005	0.358±0.039
Scene	0.305±0.022	0.677±0.028	0.281±0.091	0.439±0.115	**0.243±0.003**	0.291±0.030
Yeast	0.239±0.028	0.228±0.030	**0.216±0.022**	0.250±0.024	0.240±0.000	0.231±0.028
Bibtex	0.624±0.024	-	0.567±0.014	0.776±0.014	0.386±0.000	**0.374±0.021**
Coverage ↓						
Emotions	0.473±0.031	0.489±0.043	0.357±0.044	0.389±0.051	0.444±0.008	**0.333±0.039**
Medical	0.092±0.024	0.163±0.018	0.125±0.032	0.134±0.028	0.056±0.001	**0.049±0.014**
Image	0.248±0.015	**0.217±0.016**	0.228±0.059	0.292±0.085	0.229±0.004	0.223±0.011
Scene	0.136±0.009	0.337±0.018	0.113±0.047	0.224±0.048	0.111±0.001	**0.108±0.011**
Yeast	0.471±0.017	**0.446±0.016**	0.468±0.017	0.469±0.014	0.529±0.001	0.507±0.016
Bibtex	0.369±0.014	-	0.469±0.010	0.482±0.008	**0.255±0.000**	0.268±0.009
Average Precision ↑						
Emotions	0.619±0.027	0.567±0.028	0.731±0.024	0.704±0.034	0.674±0.010	**0.767±0.039**
Medical	0.768±0.027	0.409±0.035	0.735±0.046	0.588±0.055	0.860±0.002	**0.873±0.028**
Image	0.730±0.015	**0.771±0.017**	0.744±0.075	0.691±0.096	0.766±0.002	0.761±0.021
Scene	0.798±0.134	0.528±0.020	0.803±0.069	0.699±0.072	**0.843±0.001**	0.820±0.018
Yeast	0.754±0.017	**0.761±0.018**	0.753±0.012	0.741±0.014	0.721±0.000	0.742±0.016
Bibtex	0.312±0.016	-	0.298±0.011	0.183±0.010	0.509±0.000	**0.520±0.012**

randomly adding false positive labels under various configurations of variable p and r.

Specifically, $p \in (0,1)$ denotes the percentage of instances in the dataset that are partially tagged, and $r \in N$ represents the number of false positive labels appearing in the candidate labels. Table 1 illustrates the properties of partial multi-label datasets, such as the number of instances N(E), the number of features N(F), the number of label N(L), the average number of labels or per instances Avg(L).

Furthermore, two well-established multi-label methods ML-KNN [10], RankSVM [2] are used as the comparing algorithms. Besides, three well-established partial multi-label learning methods are also employed as the comparing approaches that consider the disambiguation for candidate labels, such as PARTICE-VLS, PARTICE-MAP [9] and DRAMA [7]. The corresponding parameters are presented in the respective literature. In this paper, k=10, g=6, θ_1= 0.5, θ_2= 0.9 and γ = 1 are empirically set.

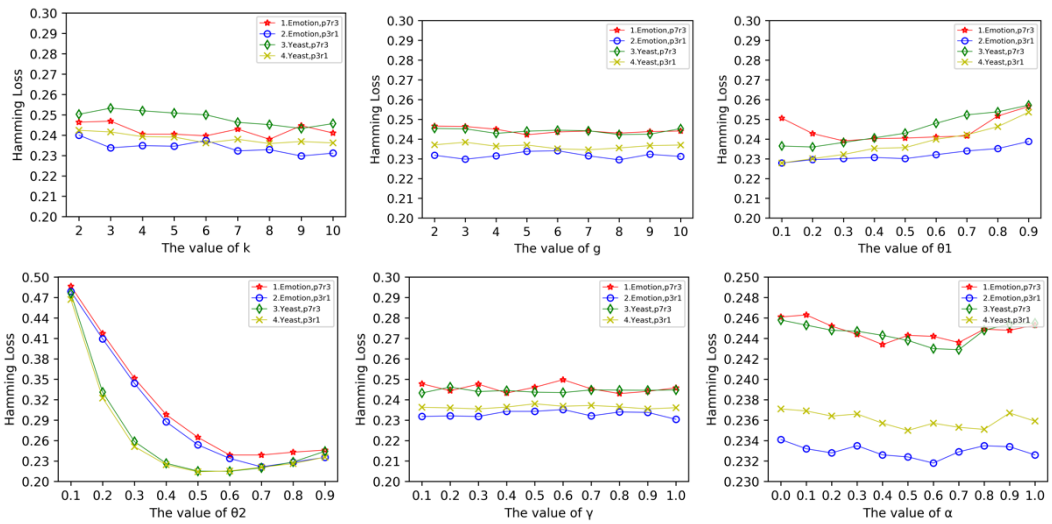

Figure 3: The performance of GLOMAN in terms of Hamming Loss with k (nearest neighbour), g (the number of clusters), θ_1 (the threshold for disambiguation), θ_2(the threshold for prediction), γ (the regularization coefficient), and α (the trade-off parameter)

We use five popular evaluation metrics for multi-label classification, including Hamming Loss, Ranking Loss, One Error, Coverage and Average Precision, whose specific meaning can be found in [3]. Specifically, for the first four metrics, "↓" indicates the smaller the better. For the average precision, "↑" indicates the larger the better. Furthermore, we employ ten-fold cross-validation to evaluate GLOMAN and above approaches. The mean and standard deviation of five metrics are reported in Table 2, where the best results are shown in bold.

3.2 Experimental Results

3.2.1 Comparison Results

Due to page constraints, we only report a portion of the experimental results on PML datasets in Table 2, where the most challenging parameters are configured with p= 7 and r =3. From the experimental result, we can conclude that:

- Out of 270 statistical tests (6 datasets × 6 methods × 5 metrics), GLOMAN separately outperforms the MLL and PML methods in most cases.
- For small datasets, such as on synthetic datasets Emotions, Medical, GLOMAN significantly outperforms other methods on all metrics.
- For large-scale datasets, such as Bibtex, GLOMAN performs better than other PML-based approaches on most evaluation metrics. Besides, GMOMAN achieves the relatively great results on each dataset, which fully proves that the method has great robustness.

3.2.2 Sensitivity Analysis

To investigate the influence of different parameters of GLOMAN, an illustrative example given by Figure 3 shows how the performance of GLOMAN (in terms of Hamming Loss) varies as each parameter changes with other parameter fixed on the Emotion and Yeast datasets. Here, when the value of one parameter various, the values for the other parameters are fixed as default parameters. According to the experimental results, it is shown that in most cases:

- The performance of GLOMAN is relatively stable under varying k, g and γ.
- The parameter θ_1 usually follow the optimal configuration in [0.2,0.5].
- The performance of GLOMAN gradually improves as $\theta_2$grows from [0.1,0.5] and becomes relatively stable in [0.6,0.9].
- To fully test the usefulness of the joint feature manifold, we further investigate the influence of α. The performance of GLOMAN improves as α in [0.4,0.7], which validates the effectiveness of joint feature manfiold structure learned by GLOMAN for solving PML problem.

4 CONCLUSIONS

By exploiting the global and local label correlations, this paper proposes a novel PML framework called GLOMAN. Firstly, the joint feature manifolds are fused by the global and local feature manifold. Then, GLOMAN attempts to disambiguate the candidate labels by utilizing the joint feature manifold structure in the label space. Experimental results on six datasets demonstrate the superiority and the effectiveness of our model.

REFERENCES

[1] Chapelle, O., Scholkopf, B., & Zien, A. (2009). Semi-supervised learning. IEEE Transactions on Neural Networks, 20(3), 542-542.

[2] Elisseeff, A., & Weston, J. (2001). A kernel method for multi-labelled classification. Advances in neural information processing systems, 14.

[3] Gibaja, E., & Ventura, S. (2015). A tutorial on multilabel learning. ACM Computing Surveys (CSUR), 47(3), 1-38.

[4] Hou, P., Geng, X., & Zhang, M. L. (2016, February). Multi-label manifold learning. In Proceedings of the AAAI Conference on Artificial Intelligence (Vol. 30, No. 1).

[5] Huang, S. J., & Zhou, Z. H. (2012). Multi-label learning by exploiting label correlations locally. In Proceedings of the AAAI Conference on Artificial Intelligence (Vol. 26, No. 1, pp. 949-955).

[6] Lyu, G., Feng, S., & Li, Y. (2020, August). Partial multi-label learning via probabilistic graph matching mechanism. In Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining (pp. 105-113).

[7] Wang, H., Liu, W., Zhao, Y., Zhang, C., Hu, T., & Chen, G. (2019, January). Discriminative and Correlative Partial Multi-Label Learning. In IJCAI (pp. 3691-3697).

[8] Xie, M. K., & Huang, S. J. (2018, April). Partial multi-label learning. In Proceedings of the AAAI Conference on Artificial Intelligence (Vol. 32, No. 1).

[9] Zhang, M. L., & Fang, J. P. (2020). Partial multi-label learning via credible label elicitation. IEEE Transactions on Pattern Analysis and Machine Intelligence, 43(10), 3587-3599.

[10] Zhang, M. L., & Zhou, Z. H. (2007). ML-KNN: A lazy learning approach to multi-label learning. Pattern recognition, 40(7), 2038-2048.

[11] Zhang, M. L., & Zhou, Z. H. (2013). A review on multi-label learning algorithms. IEEE transactions on knowledge and data engineering, 26(8), 1819-1837.

[12] Zhang, M. L., Zhou, B. B., & Liu, X. Y. (2016, August). Partial label learning via feature-aware disambiguation. In Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining (pp. 1335-1344).

[13] Zhou, Z. H. (2018). A brief introduction to weakly supervised learning. National science review, 5(1), 44-53.