A NOVEL GRAPH BASED LABEL PROPAGATION METHOD FOR HYPERSPECTRAL REMOTE SENSING DATA CLASSIFICATION

Xiaopan Wang , *Yan Hu*, *Shaojia Zhang*

Chongqing Geomatics Center, Chongqing, 401120, China

ABSTRACT

For hyperspectral image classification, we present a novel graph based semi-supervised classification method that learns from similarity and dissimilarity on labeled and unlabeled data, which contain both the adjacency graph and the dissimilar graph. Since manifold learning approach is capable of exploring the manifold geometry of data, it is suitable for calculating the adjacency graph with label similarity. A manifold learning method was utilized to calculate the adjacency graph. Dissimilarity among examples probably be used to construct the dissimilar graph, which is hard to grasp. The dissimilar probability was proposed to construct the dissimilar graph, which has effectively improved the classification accuracy of hyperspectral data in experiment.

Index Terms—*Graph; Label Propagation; Semi-Supervised Classification; Hyperspectral Remote Sensing*

1. INTRODUCTION

The graph based label propagation (LP) method[1][2], one of the widely used semi-supervised method, has been successfully applied to data classification by numerous researchers[3][4]. The most common method such as Mincut[5], Gaussian Random and Harmonic Functions (GRHF)[6], Local and Global Consistency LGC)[7], Linear Neighbor Propagation (LNP)[8], and Non-negative Sparse Representation based Label Propagation (NSRLP)[9]. Those semi-supervised learning algorithm that utilizes both labeled and unlabeled data is widely employed to solve the small size sample problem.

The graph Laplacian matrix in graph based label propagation method is obtained by constructing the data adjacency graph and choosing graph edge weights. Graph based semi-supervised classification algorithm usually applies some strategy to optimize the graph. However, we proposed a novel graph in label propagation framework for hypersepctral remote sensing data classification, which unite both the adjacency graph and the dissimilar graph. Since manifold learning approach is capable of exploring the manifold geometry of data, it is suitable for calculating the graph Laplacian in LP. The laplacian eigenmaps(LE)[10] was used to construct the adjacency graph in this paper. The dissimilar probability of each unlabeled point can be calculated by solving an l_1 optimization problem, which has a significant influence to construct the dissimilar graph. In this study, the adjacency graph and the dissimilar graph are liner combine in label propagation framework. Experiments on real hyperspetral data sets demonstrate the effectiveness of our approach.

The rest of the paper is outlined as follows. Section 2 reviews the framework of label propagation, and the way to construct the adjacency graph and dissimilar graph. Section 3 shows the experimental results, where four methods are contrast on real hypersprctral data. Finally, conclusions are summarized in section 4.

2. METHODOLOGY

2.1. The Label Propagation Framework

Under the regularization framework, the graph based label propagation is used exploit the geometry of the marginal distribution. Let $\mathbf{X}_l$=[x_1,…,x_l] denote l labeled data with labels $\{y_i\}_{i=1}^{l} \in \{1,...C\}$ and $\mathbf{X}_u$=[x_{l+1},…,x_{l+u}] denote u unlabeled data. The regularized function to be minimized is defined as:

$$f^* = \min_{f \in H_\kappa} \frac{1}{l} \sum_{i=1}^{l} V(x_i, y_i, f) + \gamma_I \|f\|^2 \tag{1}$$

where V is some loss function, $\|f\|^2$ is the manifold regularization term that reflects the smoothness of $\boldsymbol{f}$ on the data manifold. γ_I is the corresponding regularization parameters. $\|f\|^2 = \frac{1}{(l+u)^2} \sum_{i,j=1}^{l+u} \boldsymbol{f}^{\mathrm{T}} \boldsymbol{L} \boldsymbol{f}$, where **L=D-W** is the graph Laplacian matrix, $\tilde{\mathbf{L}} = \mathbf{I} - \mathbf{D}^{-1/2}\mathbf{W}\mathbf{D}^{-1/2}$ is normalized graph Laplacian. **W** is the edge weights of graph, **D** is the diagonal degree matrix of **W** given by $D_{ii} = \sum_{j=1}^{l+u} W_{ij}$. The normalizing coefficient $1/(l+u)^2$ is the nature scale factor for the empirical estimate of Laplace operator, and $\mathbf{f} = [f(x_1),..., f(x_{l+u})]^{\mathrm{T}} = \mathbf{K}\beta + b$, where b is the bias term.

2.2. Construct the Dissimilar Graph

In this paper, we attempt to build a novel graph combine with the adjacency graph and the dissimilar graph. The geometry of data is modeled with the couple graph where nodes consist of both labeled and unlabeled data points connected by edge weights. The main question is to find the dissimilar data pairs and calculate the edge weight. We try to choose the dissimilar data pairs, by solving an l_1 optimization problem on sparse representation (SR). The steps of constructing the dissimilar graph are as follows:

Firstly, calculate the class-probability of each unlabeled data by. $P((\boldsymbol{X}_0)_i)=\hat{\alpha}^{\mathrm{T}} * \boldsymbol{Y}_L$, where $(\mathbf{X}_0)_i$ is an unlabeled data , $\boldsymbol{Y}_L \in R^{l\times c}$ is the true label of train data. $\hat{\alpha}$ is the sparse coefficient vector, by solving the l_1 optimization problem:

$$\hat{\alpha}=\arg\min_{\alpha}\{\|(\boldsymbol{X}_0)_i\text{-}\boldsymbol{A}\alpha\|_2^2+\lambda\|\alpha\|_1\} \tag{2}$$

Secondly, utilize the label-probability to find the dissimilar data pairs. Let $\mathrm{P}(x_i, x_j)$ denotes the probability of x_i and x_j that have the same label. Through the total probability formula, $\mathrm{P}(x_i, x_j)$ can be solved by:

$$P(x_i, x_j) = \sum_{c=1}^{C} P(x_{c,j})P(x_i/x_{c,j}) \ = \hat{P}(x_i)^T \hat{P}(x_j) \tag{3}$$

Where x_i and x_j both unlabeled, or one labeled and the other unlabeled label. $\mathrm{P}(x_i, x_j)$=0 while x_i and x_j have different label, $\mathrm{P}(x_i, x_j)$=1 if x_i and x_j have same label.

Let **T** as an empirical threshold coefficient. $\mathrm{P}(x_i, x_j) \leq \mathrm{T}$ denote x_i and x_j may have different labels,which will be linked in the dissimilar graph. On the otherwise, if $\mathrm{P}(x_i, x_j) > \mathrm{T}$, x_i and x_j cannot be linked in dissimilar graph.

Finally, calculate the "dissimilar weight" of dissimilar graph. if $\mathrm{P}(x_i, x_j) > \mathrm{T}$, $W_{ij} = 0$. If $\mathrm{P}(x_i, x_j) \leq \mathrm{T}$, x_i and x_j are "dissimilar data-pairs", let weights W_{ij} equals to $\exp(-\|x_i - x_j\|^2/2\delta^2$. However, the regularization term of label propagation framework can be rewritten as:

$$\sum_{i,j=1}^{l+u} \boldsymbol{f}^{\mathrm{T}}\tilde{\boldsymbol{L}}\boldsymbol{f} = \sum_{i,j=1}^{l+u} \tilde{W}_{i,j}(f(x_i)+f(x_j))^2 \tag{4}$$

Where $\tilde{\mathbf{L}} = \tilde{\mathbf{D}} + \tilde{\mathbf{W}}$ is the, given by $\tilde{\mathbf{W}}$ is the edge weight of the dissimilar graph.

2.2. The Combined Graph Based Label Propagation Method

Linear combination is a simple and effective way to unite both similar and dissimilar graph Laplacian. Let M denote the mixed graph Laplacian matrix, $\mathbf{M} = (1-\alpha)\mathbf{L} + \alpha\tilde{\mathbf{L}}$ where a is the scale factor for the empirical estimate of dissimilarity and similarity. $\tilde{\mathbf{L}}$is the dissimilar graph, **L=D-W** is the adjacency graph Laplacian, given by:

$$\sum_{i,j=1}^{l+u} \boldsymbol{f}^{\mathrm{T}}\boldsymbol{L}\boldsymbol{f} = \sum_{i,j=1}^{l+u} W_{i,j}(f(x_i)-f(x_j))^2 \tag{5}$$

In this paper, we construct the adjacency graph with labeled and unlabeled data by using k-nearest-neighbors, where spectral information divergence (SID) was utilized to choose neighbors of the adjacency graph. Then calculate the edge weight matrix **W** by using LE.

3. EXPERIMENTAL

3.1. Data Description Analysis of DSCLP

The experiment hyperspectral data was collected by the Hyperion scanner on the EO-1 satellite, which has a 30-m spatial resolution, covering the 357-2576nm of the spectrum in 10-nm bands, over Okavango Delta Botswana (BOT) in May 2001. After removing the un-calibrated and noisy bands, 149-bands are remained for BOT respectively.

In the BOT data, labeled data consist of nine identified land cover types in seasonal swamps, occasional swamps, and woodlands. Class 3(Riparian) and class 6 (Woodlands) are very alike among the total 9 classes. They have mixed spectral signatures with subtle differences and are very difficult to classify. Focusing on the classification of these difficultly distinguished classes, we also choose class 3 (237 points) and class 6(199 points) of BOT data, as our experiment data sets. All the points will be dividing into two subsets, one for training and other for testing

3.2. Analysis of DSCLP

In the proposed dissimilarity and similarity combining label propagation (DSCLP) method, we choose two methods to calculate the dissimilar weights. One is the heat kernels Where $W(x_i, x_j) = \exp(-\|x_i - x_j\|^2/2\delta^2)$, δ is the kernel width, and varied in the range $\{0.001, 0.01, 0.1, 1, 10\}$. The other is binary weight, where $\tilde{W}(x_i, x_j) \in \{0,1\}$. For the other four three parameters (k, α, and T): Parameters k which is the number of nearest neighbors in adjacency graph was varied in the range {5,...,50} with step of 5; T is an empirical threshold of dissimilar probability, which varied in the range (0.1,0.9) with step of 0.1; α appoint the tradeoff between the similar graph and the dissimilar graph, which varied in the range (0.1,1) with step of 0.1. It should be note that, if $\alpha = 0$ DSCLP equals to GRHF method.

Four classifiers were applied to the data sets. There are GRHF, LNP, NSRLP and DSCLP. The dissimilar graph with binary weight in DSCLP is contrast. It should be note that the k-NN was employed as the similarity measurement to search neighbors and use radial basis function to calculate graph weigh for the first two label propagation method. NSRLP also is semi-supervised label propagation, which use SR to calculate the graph Laplacian. We randomly selected labeled data with number of {3, 5, 10, 20, 30} for each class.

The average classification accuracies of 10 replications are obtained, where the results of BOT(C1-C9)data with the optimal parameter combination chosen by exhaustive search are shown in Fig.1 and Table 1, Table 2. The results of Class 3 and Class 6 of BOT data are shown in Table 3. It should be note that, DSCLP(Binary) denote the weight of dissimilar graph is calculated by binary weight, and DSCLP(Heat) denote the weight of dissimilar graph is calculated by heat kernel.

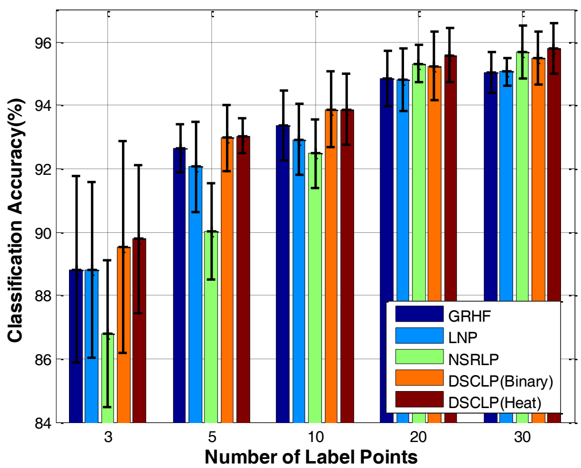

Fig.1 Overall Classification Accuracy results by different methods over BOT(C1-C9) data

Table 1 Omission error of each class over BOT data

Method / class	Omission Error, GRHF	LNP	NSRLP	DSCLP (Binary)	DSCLP (Heat)
C1	0	1.47	0.18	1.74	0
C2	6.33	3.86	12.09	4.87	3.99
C3	20.55	20.73	14.57	18.17	17.56
C4	2.93	5.2	3.33	2.85	2.36
C5	28.97	22.54	19.6	25.87	25.71
C6	18.1	15.26	30.15	18.61	17.01
C7	1.98	4.68	10	2.34	3.87
C8	1.06	5.06	5.18	1.06	3.76
C9	13.29	19.74	19.87	12.5	13.29
Total	10.36	10.95	12.77	9.78	9.73

Fig.1 shows the classification accuracies of BOT(C1-C9) data set. Table 1 shows the omission error and table 2 shows the commission error, for each class with 3 labeled points by using the above methods. Table 3 shows the classification accuracy over Class 3 and Class 6 of BOT data. Several observations can be obtained: 1) DSCLP(Heat) performs similar to DSCLP(Binary). 2) DSCLP(Heat) always produced higher classification accuracies than other method. 3) Both DSCLP(Heat) and DSCLP(Binary) produce slightly better performance than GRHF and LNP, which prove that the dissimilar graph improve the classification accuracy for BOT data. 4) With the number of labels decreases, DSCLP have a better performance for BOT(C1-C9)data, especially contrast with NSRLP.

Table 2 Commission error of each class over BOT data

method / class	Commission Error GRHF	LNP	NSRLP	DSCLP (Binary)	DSCLP (Heat)
C1	0.91	3.3	2.48	1.35	0.45
C2	1.3	5.91	8.16	2.96	3.35
C3	8.27	8.89	16.71	9.13	10.19
C4	0	2.66	5.37	0	0
C5	10.52	10.86	14.71	10.77	11.02
C6	26.04	22.07	14.19	23.3	21.77
C7	13.33	17.69	16.85	9.18	15.54
C8	18.12	13.24	22.94	16.51	12.2
C9	15.63	8.34	17.74	15.51	10.99
Total	10.46	10.33	13.24	9.86	9.5

Table 3 Overall Accuracy of Class 3 and Class 6 of BOT data

Method / Label points	GRHF	LNP	NSRLP	DSCLP (Binary)	DSCLP (Heat)
3	85.45	85.02	85.91	88.21	88.21
5	86.34	86.14	88.86	89.66	89.66
10	87.63	86.36	90.24	91.24	91.34
20	89.46	89.78	93.18	94.48	94.95
30	90.49	89.85	94.45	94.49	95.13

4. CONCLUSION

This paper proposed a novel graph Laplacian, which unite both the adjacency graph and the dissimilar graph in label propagation framework. For the adjacency graph, *k*-nearest-neighbors used to choose neighbors, and LE used to calculate the graph weights. For the dissimilar graph, dissimilar data pairs were chosen by solving an l_1 optimization problem on SR. Contrast with three label propagation methods for hyperspectral image classification. From the experiments on real hyperspectral data sets, we conclude that the DSCLP has a better performance than other label propagation classifiers.

It should be note that multi-graph by other characteristics also can be unite in this framework. However, the combination of similar graph and adjacency graph is not only limited to linear method. Combine multi-graph with nonlinear method will be in the future work.

Acknowledgments: This work was supported by Science and Technology Project of Ministry of Housing and Urban-Rural Development of China (Grant No.2017-K8-046)

5. REFERENCES

[1] Zhu, X. "Semi-supervised Learning with Graphs," Doc Thesis Carnegie Mellon Univ., vol. 6395, no. 10, pp. 2465-2472, 2005.

[2]Belkin M., Niyogi P., and Sindhwani V., “Manifold regularization: A geometric framework for learning from labeled and unlabeled examples,” *JMLR.org,* 2004.

[3]Bai J., Xiang S., and Pan C., “A Graph-Based Classification Method for Hyperspectral Images,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 51, no. 2, pp. 803-817, 2013.

[4]Camps-Valls G., Bandos Marsheva T., ans Zhou D., “Semi-supervised graph-based hyperspectral image classification,” Geoscience and Remote Sensing, *IEEE Transactions on Geoscience and Remote Sensing*, vol. 45, no. 10, pp. 3044-3054, 2007.

[5]Blum A., Chawla S., “Learning from labeled and unlabeled data using graph mincuts,” *Processes 18th international conference on Machine Learning*, vol. 3175, pp. 19-26, 2001.

[6]Zhu X., Ghahramani Z., Lafferty J., “Semi-supervised learning using gaussian fields and harmonic functions,” *Proc ICML*. pp. 912-919, 2003.

[7]Zhou D., Bousquet O., Lal T. N., and Weston J., “Learning with Local and Global Consistency,” *International Conference on Neural Information Processing Systems MIT Press*, pp. 321-328, 2003.

[8]Wang F., Zhang C., “Label propagation through linear neighborhoods,” *IEEE Transactions on Knowledge and Data Engineering*, vol. 20, no. 1, pp. 55-67, 2008.

[9]Yang N., Sang Y., He R., and Wang, X., “Label propagation algorithm based on non-negative sparse representation,” *Journal of Dalian University of Technology*, vol. 6330, no. 2, pp. 348-357, 2010.

[10]Mikhail Belkin, Partha Niyogi, “Laplacian Eigenmaps for Dimensionality Reduction and Data Representation,” *MIT Press*, 2003.