

# COUPLE GRAPH BASED LABEL PROPAGATION METHOD FOR HYPERSPECTRAL REMOTE SENSING DATA CLASSIFICATION

XiaoPan WANG<sup>1</sup>, Yan Hu<sup>1</sup>, Jing Chen<sup>1</sup>

<sup>1</sup>Chongqing Geomatics Center, Chongqing, 401121, China-(xpwang, huyan, cj)@dl023.net

Commission III, WG III/4

**KEY WORDS:** Graph, Label Propagation, Semi-Supervised method, Classification, Hyperspectral Remote Sensing

## ABSTRACT:

Graph based semi-supervised classification method are widely used for hyperspectral image classification. We present a couple graph based label propagation method, which contains both the adjacency graph and the similar graph. We propose to construct the similar graph by using the similar probability, which utilize the label similarity among examples probably. The adjacency graph was utilized by a common manifold learning method, which has effective improve the classification accuracy of hyperspectral data. The experiments indicate that the couple graph Laplacian which unite both the adjacency graph and the similar graph, produce superior classification results than other manifold Learning based graph Laplacian and Sparse representation based graph Laplacian in label propagation framework.

## 1. INTRODUCTION

As semi-supervised classifiers, graph based label propagation (LP) methods experienced fast growth over the past decade(Zhu, 2005), (Camps-Valls, 2007) and (Bai, 2013). The most common method such as Mincut(Blum et al., 2001), Gaussian Random and Harmonic Functions(Zhu, et al. 2003) short by GRHF, Local and Global Consistency(Zhou, et al. 2003) short for LGC, Linear Neighbor Propagation(Wang, et al. 2007) short for LNP, and the local tangent space alignment(Zhang, et al. 2004) short for LTSA, are widely applied to data classification by numerous researchers. For hyperspectral image classification, supervised classifier required a large number of labeled data due to the high dimensional spectra. However, labeled instances are often difficultly, costly, or time consuming to obtain. Those semi-supervised learning algorithm that utilizes both labeled and unlabeled data is widely employed to solve the small size sample problem.

Usually, the graph Laplacian matrix in graph based label propagation method is obtained by constructing the data adjacency graph and choosing graph edge weights.  $k$ -nearest-neighbor method used to construct the data adjacency, and graph weight are chosen by binary weights, heat kernel weights, or Euclidean distance weights. However, these weights are only determined by the pair wise distances between data points, ignoring the neighborhood relations and thereby potentially underutilizing available information.

In this paper, we proposed a novel graph in label propagation framework which unites both the adjacency graph and the similar graph. Since manifold learning(ML) approach is capable of exploring the manifold geometry of data(Belkin, et la. 2006), it is suitable for calculating the graph Laplacian in LP. The laplacian eigenmaps(LE) which proposed by (Belkin, 2003), was used to construct the adjacency graph in this paper. The class-probability of each unlabeled point can be calculated by solving an  $l_1$  optimization problem, which has a significant influence to construct the similar graph. In this study, the adjacency graph and similar graph are liner combine in label propagation framework. Experiments on real hyperspectral 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 similar graph. Section 3 shows the experimental results, where four methods are contrast on two hypersprctral data, and two factors that influence the graph Laplain are analyzed. 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  $\mathbf{X}_l=[x_1, \dots, x_l]$  denote  $l$  labeled data with labels  $\{y_i\}_{i=1}^l \in \{1, \dots, C\}$  and  $\mathbf{X}_u=[x_{l+1}, \dots, x_{l+u}]$  denote  $u$  unlabeled data. The regularized function to be minimized is defined as:

$$f^* = \min_{f \in H_c} \frac{1}{l} \sum_{i=1}^l V(x_i, y_i, f) + \gamma_1 \|f\|^2 \quad (1)$$

where  $V$  = some loss function

$\gamma_l$  = the corresponding regularization parameters.

$\|f\|^2$  = the manifold regularization term that reflects the smoothness of  $f$  on the data manifold, which defined as:

$$\|f\|^2 = \frac{1}{(l+u)^2} \sum_{i,j=1}^{l+u} f^T M f \quad (2)$$

where

$\mathbf{M} = \mathbf{D} - \mathbf{W}$  is the graph Laplacian matrix

$\mathbf{W}$  = the edge weights of graph

$\mathbf{D}$  = the diagonal degree matrix of  $\mathbf{W}$  given by  $D_{ii} = \sum_{j=1}^{l+u} W_{ij}$ .

$1/(l+u)^2$  = the normalizing coefficient, the nature scale factor for the empirical estimate of Laplace operator.

It should be note that,  $\mathbf{f} = [f(x_1), \dots, f(x_{l+u})]^T = \mathbf{K}\beta + b$ , where  $b$  is the bias term.  $\tilde{\mathbf{M}} = \mathbf{I} - \mathbf{D}^{-1/2}\mathbf{W}\mathbf{D}^{-1/2}$  is normalized graph Laplacian.

### 2.2 The couple graph based label propagation method

In this paper, we attempt to build a couple graph combine with the adjacency graph and the similar 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 couple graph Laplacian matrix is defined as follows:

$$\mathbf{M} = (1 - \alpha)\mathbf{L} + \alpha\tilde{\mathbf{L}} \quad (3)$$

Where  $\mathbf{M}$  = the couple graph Laplacian matrix,

$\mathbf{L}$  = the adjacency graph matrix,

$\tilde{\mathbf{L}}$  = the similar matrix,  $\alpha$  is the scale factor for the empirical estimate of adjacency graph and similar graph.

$\alpha$  = the tradeoff between the similar graph and the adjacency graph.

$\mathbf{L} = \mathbf{D} - \mathbf{W}$  is the adjacency graph Laplacian, given by  $\sum_{i,j=1}^{l+u} \mathbf{f}^T \mathbf{L} \mathbf{f} = \sum_{i,j=1}^{l+u} W_{i,j} (f(x_i) - f(x_j))^2$  Construct the adjacency graph

with labeled and unlabeled data is using  $k$ -nearest-neighbors, where spectral information divergence (SID) was utilized to choose neighbors of the adjacency graph. Then calculate the edge weight matrix  $\mathbf{W}$  by using LE.

$\tilde{\mathbf{L}} = \tilde{\mathbf{D}} - \tilde{\mathbf{W}}$  is the similarity graph Laplacian, that  $\sum_{i,j=1}^{l+u} \mathbf{f}^T \tilde{\mathbf{L}} \mathbf{f} = \sum_{i,j=1}^{l+u} \tilde{W}_{i,j} (f(x_i) + f(x_j))^2$ . where  $\tilde{\mathbf{W}}$  is the edge weight of the similar graph. The main question is to find the similar data pairs and calculate the edge weight. We try to choose the similar data pairs, by solving an  $l_1$  optimization problem on sparse representation (SR):

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

Secondly, utilize the class-probability to find the similar data pairs. Calculate the class-probability by:

$$P(x_i, x_j) = \sum_{c=1}^C P(x_{c,j}) P(x_i / x_{c,j}) = \hat{P}(x_i)^T \hat{P}(x_j) \quad (4)$$

where

$x_i, x_j$  = unlabeled, or one labeled and the other unlabeled label

$P(x_i, x_j)$  = the probability of  $x_i$  and  $x_j$  that have the same label.

Set  $T$  as an empirical threshold, when  $P(x_i, x_j) > T$  denote  $x_i$  and  $x_j$  may have the same label, and set the weights of  $x_i$  and  $x_j$  as  $W_{ij} = \exp(-\|x_i - x_j\|^2 / 2\delta^2)$ . On the otherwise, if  $P(x_i, x_j) \leq T$  denotes  $x_i$  and  $x_j$  may have the different labels, set  $W_{ij} = 0$ .

## 3. EXPERIMENTAL

### 3.1 Data Description

Two hyperspectral data were used for the experiment. The first one 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. The second is the 224-Band AVIRIS data set, which was collected over the Kennedy Space Center(KSC) in March 1996, KSC has a 18-m spatial resolution and a 10-nm spectral resolution over the range of 400-2500nm. After removing the un-calibrated and noisy bands, 149-bands and 176-bands are remained for BOT and KSC data respectively.

In the BOT data, labeled data consist of nine identified land cover types, and classes 3(Riparian) and classes 6 (Woodlands) are very alike among the total 9 classes. In the KSC data, 13 land cover classes were labeled. The classes of Cabbage Palm/Oak Hammock (classes 4) and Slash Pine (classes 5) are all trees that grow in upland; they have mixed spectral signatures with subtle differences and are very difficult to classify. Focusing on the classification of the novel graph Laplacian, we choose classes 3 and classes 6 of BOT data, classes 4 and classes 5 of KSC data as our experiment data sets. "One Versus Other" classification strategy can be used in the multiclass classifier. All the points will be dividing into two subsets, one for training and other for testing.

### 3.2 Analysis of CGLP

In the proposed couple graph based label propagation (CGLP) method, we chose radial basis function (RBF) kernels,  $K(x_i, x_j) = \exp(-\|x_i - x_j\|^2 / 2\delta^2)$ , where  $\delta$  is the kernel width, and varied in the range  $\{0.001, 0.01, 0.1\}$ . For the other three parameters ( $k, \alpha$ , and  $T$ ): Parameters  $k$  which is the number of nearest neighbors in adjacency graph was varied in the range  $\{5, \dots, 20\}$  with step of 5;  $T$  is an empirical threshold of similar probability, which varied in the range  $(0.1, 0.9)$  with step of 0.1;  $\alpha$  appoint the tradeoff between the similar graph and the adjacency graph, which varied in the range  $(0, 1)$  with step of 0.1.

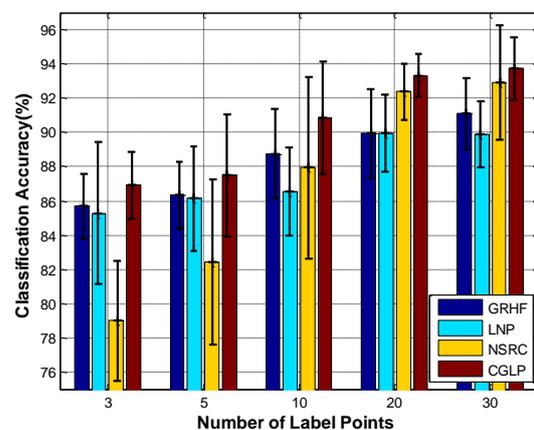


Figure 1 Classification results by different methods over Classes 3 and Classes 6 of BOT data

Four classifiers were applied on the two data sets. There are GRHF, LNP, NSRC(Non-negative Sparse Representation Classifier) and CGLP. NSRC is sparse representation based supervised method. It should be note that the k-NN method was employed as the adjacency measurement to search neighbors

and use radial basis function to calculate graph weigh for the first two label propagation method. We randomly selected labeled data with number of {3, 5, 10, 20, 30} for each class. The average classification accuracies of 20 replications are obtained, where the results with the optimal parameter combination chosen by exhaustive search are shown in figure 1 and figure 2.

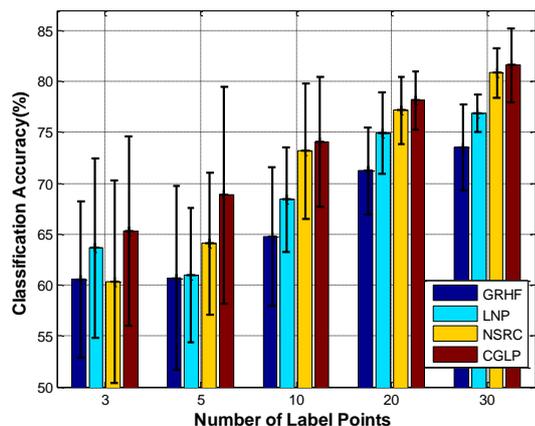


Figure 2 Classification results by different methods over Classes 4 and Classes 5 of KSC data

Figure 1 shows the overall accuracies of Classes 3 and Classes 6 of BOT data set and figure 2 shows the overall accuracies of Classes 4 and Classes 5 of KSC data set, by using the above four method. Several observations can be obtained: 1) CGLP always produced higher classification accuracies than other four methods. For KSC data, CGLP produce slightly better performance than other three methods. For BOT data, CGLP produce significantly outperform than other three methods. 2) With the number of labels decreases, CGLP have a better performance than NSRC. 3) Both KSC and BOT data, the NSRC improved fast by enhance the number of label data.

### 3.3 Analysis of threshold T and coefficient $\alpha$

To analyze the impact of the empirical threshold T to graph Laplacian, we selected 30 labeled samples for each class, and changed the parameter T with the steps of 0.1 in the range {0.3, ..., 0.7},  $\delta$  and  $k$  as shown above. It can be seen that the value of T can not be too big or too small. The T value is meaningless if T is too small. Too big, the similar graph has little contribution to graph Laplacian. Through the adjustment of T, and seek out those important and meaningful points. The random test data for Classes 4 and Classes 5 of KSC data were presented in figure 3.

To analyze the impact of tradeoff between the similar graph and the adjacency graph. We fixed the empirical threshold T fixed to 0.5, and selected 30 labeled samples for each class,  $\delta$  and  $k$  as shown above. According to experience, the value of  $\alpha$  better less than 0.5. We changed the parameter  $\alpha$  with the steps of 0.1 in the range {0, ..., 0.4}. With the increase of  $\alpha$ , the classification accuracy is increased. The random test data experimental results of Classes 4 and Classes 5 of KSC data were presented in figure 4.

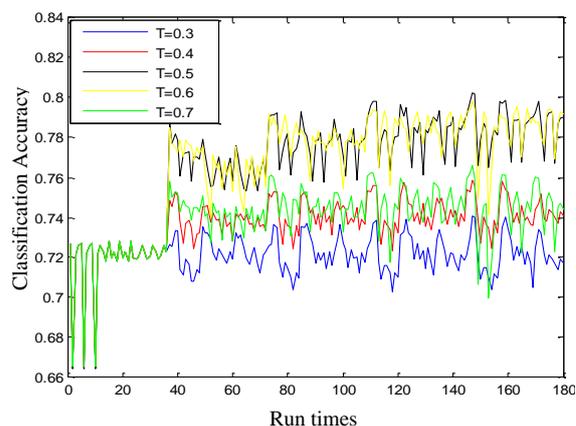


Figure 3 the impact of T for Classes 4 and 5 of KSC data

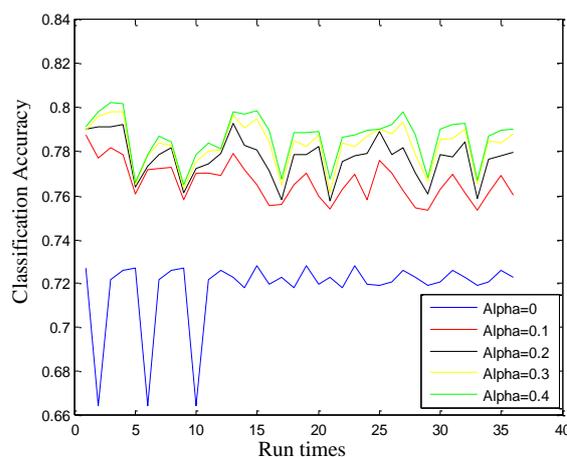


Figure 4 the impact of  $\alpha$  or Classes 4 and 5 of KSC data

## 4. CONCLUSION

This paper proposed a couple graph Laplacian, which unite both the adjacency graph and the similar 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 similar graph, similar data pairs were chosen by solving an  $l_1$  optimization problem on SR. Contrast with two label propagation methods, there are GRHF, LNP, and NSRC, for hyperspectral image classification. From the experiments on two hyperspectral data sets, we conclude that the CGLP has a better performance than other three classifiers. Moreover, when the number of label point is small, CGLP achieve superior accuracies. In addition, NSRC improved fast by enhance the number of label data, especially for BOT data.

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.

## ACKNOWLEDGEMENTS

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

## REFERENCES

Zhu, X., Ghahramani, Z., Mit, T. J. 2005. Semi-Supervised Learning with Graphs. *International Joint Conference on Natural Language Processing*, Vol.6493, pp.2465 - 2472.

Bai, J., Xiang, S., Pan, C., 2013. A Graph-Based Classification Method for Hyperspectral Images. *IEEE Transactions on Geoscience and Remote Sensing*, 51(2), pp. 803-817.

Camps-Valls. G., Marsheva T. V. B., Zhou, D., 2007. Semi-supervised graph-based hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing*, 45(10), pp. 3044-3054.

Blum, A., Chawla, S., 2001. Learning from labeled and unlabeled data using graph mincuts. *International Conference on Machine Learning*, Vol.3175, pp. 19-26.

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

Zhou, D., Bousquet, O., Lal, T. N., Weston, J., 2003. Learning with local and global consistency. *International Conference on Neural Information Processing Systems*, Vol.16, pp. 321-328).

Wang, F., Zhang, C., 2007. Label propagation through linear neighborhoods. *IEEE Transactions on Knowledge and Data Engineering*, 20(1), pp. 55-67.

Zhang, Z., Zha, H., 2005. Principal Manifolds and Nonlinear Dimensionality Reduction via Tangent Space Alignment. *Society for Industrial and Applied Mathematics*, Vol.26, no.1, pp. 313-338, Dec.2004.

Belkin, M., Niyogi, P., Sindhwani, V., 2006. Manifold regularization: A geometric framework for learning from labeled and unlabeled examples. *The Journal of Machine Learning Research*, pp. 2399-2434.

Belkin, M., Niyogi, P., 2003. Laplacian Eigenmaps for dimensionality reduction and data representation. *Neural Computation*, Vol.15, no.6, pp. 1373-1396.

Yang, N., Sang, Y., He, R., 2010. Label propagation algorithm based on non-negative sparse representation. *Life System Modeling and Intelligent Computing*. Springer Berlin Heidelberg, pp. 348-357.