SEMI-SUPERVISED HYPERSPECTRAL BAND SELECTION VIA SPARSE LINEAR REGRESSION AND HYPERGRAPH MODELS

Zhouxiao Guo[1]*, Haichuan Yang*[1]*, Xiao Bai*[1]*, Zhihong Zhang*[2]*, Jun Zhou*[3]

1. School of Computer Science and Engineering
Beihang University, Haidian District, Beijing, China
2. Xiamen University, Siming District, Xiamen, China
3. School of Information and Communication Technology
Griffith University, Nathan, QLD 4111, Australia

ABSTRACT

Band selection is an important step towards effective and efficient object classification in hyperspectral imagery. In this paper, we propose a semi-supervised learning method for band selection based on a sparse linear regression model. This model uses a least absolute shrinkage and selection operator to compute the regression coefficients from both labeled and unlabeled samples. These coefficients are then used to compute a contribution score for each band, which allows bands with high scores being selected for the testing step. During this process, unlabeled samples also contribute to the coefficients calculation. In order to propagate the labels to these samples, a hypergraph is first built to describe the relationship between labeled and unlabeled samples. This leads to an adjacency matrix whose entries are the sum of corresponding weights of hyperedges. Then matrix subspace learning method is used to estimate the labels of unlabeled samples. The proposed method is evaluated on the APHI dataset. Comparison with several baseline methods has shown the advantages of the proposed method on the pixel-level classification.

Index Terms— semi-supervised learning, band selection, hyperspectral image, sparse linear regression, hypergraph

1. INTRODUCTION

Hyperspectral sensors can generate remote sensing images in high spectral resolution with hundreds of contiguous spectral bands. Such spectral information is closely related to the material property of targets, which makes hyperspectral imagery very suitable for land cover classification, object detection, and mining applications [1].

One problem with hyperspectral data is that neighboring bands are often highly correlated and contain redundant information. This has greatly reduced the efficiency of image processing and analysis. To address this problem, many band selection methods have been reported, which extract those bands that contain highly discriminative information for pixel-level classification [2]. Representative work include constrained band selection [3], mutual information [4], and affinity propagation [5]. These methods either adopt different criteria to rank spectral bands, or divide bands into groups and then select those most representative ones from each group. Of particular interest is the work of Du et al [6], which uses N-FINDR and linear prediction to perform an initial band selection, and then refines the band subset using collaborative sparse model. This has motivated us to use sparse linear regression to solve the band selection problem, which produces a sparse vector of coefficients corresponding to the input variables, so that only part of the bands are used for discriminating objects in the image.

The linear regression model requires label of hyperspectral pixels be known. This has made the band selection a supervised learning process. On the other hand, most band selection methods are unsupervised due to the difficulties in getting ground truth data. In recent, semi-supervised learning approaches have been reported [7, 8] for band selection, which use adaptive affinity propagation or a combination of Fisher's discriminative analysis and Graph Laplacian to model both labeled and unlabeled data.

In this paper, we propose a semi-supervised learning approach so that unlabeled data can contribute to the band selection process in addition to a small number of labeled data. This method is based on a sparse linear regression model which uses a least absolute shrinkage and selection operator (lasso) [9] to compute the regression coefficients from both labeled and unlabeled samples. These coefficients are then used to generate a contribution score for each band, which allows bands with high scores be selected. In order to propagate the labels to unlabeled image patches, a hypergraph is first built to describe the relationship between samples, and to compute an adjacency matrix whose entries are the sum of corresponding weights of hyperedges. Then matrix subspace learning method is used to estimate the labels of unseen samples.

The rest of the paper is organized as follows. Section 2 briefly introduces the sparse linear repression model. Section

3 describes the band selection method. Experimental results are reported in Section 4, while conclusions are given in Section 5.

2. SPARSE LINEAR REGRESSION

We commence from the definition of the problem. Given a set of L hyperspectral pixels $\mathbf{X}_{tr} = \{x_1, \ldots, x_L\}$ with labels $\mathbf{Y}_{tr} = \{y_1, \ldots, y_L\}$, and a set of U pixels $\mathbf{X}_{te} = \{x_{L+1}, \ldots, x_{L+U}\}$ whose labels are unknown. x_i contains spectral data with M bands in one of the C classes. We can build the following linear classifier to characterize the relationship between input hyperspectral data and output image class information

$$\mathbf{Y} = \mathbf{X}\mathbf{w} + \varepsilon \tag{1}$$

where$\mathbf{X}$ is an $(L+U) \times M$ input matrix that contains both labeled and unlabeled data, $\mathbf{Y}$ denotes the $(L+U) \times C$ output, $\mathbf{w}$ is a projection matrix, and ε is a noise matrix. Here, Y_i takes the form of $[0, ..., 0, 1, 0, ..., 0]^T$.

If both $\mathbf{X}$ and $\mathbf{Y}$ are known, our goal is to learn the optimal $\mathbf{w}$ for band selection. This can be achieved by finding the optimal $\mathbf{w}_c$ vector for each class $c \in C$. Here we proposed to use least absolute shrinkage and selection operator (lasso) [9, 10] which poses a sparse constraint on the linear regression model:

$$\mathbf{w}_c^* = \arg\min_{\mathbf{w}_c} |\mathbf{Y}_c - \mathbf{X}\mathbf{w}_c|^2 + \lambda |\mathbf{w}_c|_1 \tag{2}$$

$$|\mathbf{w}_c|_1 = \sum_{j=1}^{m} |\mathbf{w}_{cj}| \tag{3}$$

where $\mathbf{Y}_c$ is the c^{th} column of the $\mathbf{Y}$ matrix, λ is a parameter in $(0, 1)$ that controls the contribution of the L_1 penalty term,and $\mathbf{w}_{cj}$ is the coefficient correspond to the j^{th} band of the c_{th} class.

Combining optimal weights for each class, we define the following ranking score:

$$S(j) = \max_j |\mathbf{w}_{cj}| \tag{4}$$

where j is the band index. By ranking the band-score in a descending order, the top T hyperspectral bands can be selected for the testing stage.

3. LABEL PROPAGATION

In Equation 1, we assume the labels of all input data are known. However, $\mathbf{X}$ consists of both labeled and unlabeled data. This suggests that we have to estimate the labels of the unknown data based on the labeled data in order to make the above method feasible. This is a typical label propagation problem in the semi-supervised learning setting. To solve this problem, we start from calculating an adjacency matrix using a hypergraph model.

3.1. Hypergraph Model

A hypergraph is a generalization of a graph in which an edge can connect any number of vertices [11]. Here, we build a K-uniform hypergraph for which hyperedge has the same number of K vertices. A hyperedge weight containing K vertices is computed as follows

$$W_{i_1,i_2,\ldots,i_K} = K \frac{I_{x_{i_1},x_{i_2},\ldots,x_{i_K}}}{H(x_{i_1}) + H(x_{i_2}) + \ldots + H(x_{i_K})} \tag{5}$$

where $W_{i_1,i_2,\ldots,i_K}$ is the weight of hyperedge i which contains $i_1, .., i_K$ vertices, and $I_{x_{i_1},x_{i_2},\ldots,x_{i_K}}$ is the multivariate mutual information which stores the relationships among features. $H(x_i)$ represents the Shannon entropy of a discrete random variable with possible values and probability mass function. $I_{x_{i_1},x_{i_2},\ldots,x_{i_K}}$ can be computed as

$$I_{(x_{i_1},x_{i_2},\ldots,x_{i_K})} = \sum_{k=1}^{N} (-1)^{k-1} \sum_{\substack{X \subset (x_{i_1}, x_{i_2}, \ldots, x_{i_K}) \\ |X| = k}} H(X) \tag{6}$$

where $I_{x_{i_1},x_{i_2},\ldots,x_{i_K}}$ is the information context introduced by Han [12]. It should be mentioned here that the greater the value of $W_{i_1,i_2,\ldots,i_K}$ is, the more relevant the K vectors are.

In order to deal with the relationship between two feature points using the weight of hyperedge W, we transform the K-uniform hypergraph into an adjacency matrix P. The sum of all weights $W_{i_1,..,i_K}$ containing features indexed by a and b becomes the edge weight P_{ab}, such that

$$P_{ab} = \sum_{i_1=1} \ldots \sum_{i_{K-2}=1} W_{i_1,\ldots,i_{K-2},a,b} \tag{7}$$

where P is an $(L+U) \times (L+U)$ matrix.

3.2. Subspace Learning

With the adjacency matrix ready, harmonic functions are used to provide the learning framework with consistent probabilistic semantics [13]. In order to compute the harmonic function, the adjacency matrix P is split into four blocks at the L^{th} row and U^{th}column $P = \begin{bmatrix} P_{LL} & P_{LU} \\ P_{UL} & P_{UU} \end{bmatrix}$ Let $F = \begin{bmatrix} f_L \\ f_U \end{bmatrix}$ where f_U denotes the output of the unlabeled data points. $f_L \in R^{L \times C}$, $f_{ij} = 1$ if x_i is labeled as j and $f_{ij} = 0$ otherwise. So the propagated labels of the unknown data are given by

$$f_U = (D_{UU} - P_{UU})^{-1} P_{UL} f_L = (I - P_{UU})^{-1} P_{UL} f_L \tag{8}$$

Note that the f_U stores the probabilities that an unlabeled sample belongs to each class. We have to convert it into the form of $[0, ..., 0, 1, 0, ..., 0]^T$. To do so, a projection matrix U is constructed from F so that $\mathbf{Y} = \mathbf{X}U$. Here we build a linear classifier $y = U^T x + b$ where x is a sample in the hyperspectral imagery, b is a bias term. If it turns out that y is close

to t_j where $t_j = [0, ..., 0, 1, 0, ..., 0]^T$, then x is classified as class j. At the same time, we can guarantee that the result of classification from the linear classifier is consistent with that of the above semi-supervised learning matrix F. Based on this consideration, we can define a function:

$$arg \min_{U,b} \alpha||U||^2 + \sum_{i=1}^{N}\sum_{j=1}^{C} F_{i,j}||U^T x_i + b - tj||^2 \quad (9)$$

The formula 9 is a regression problem and can be easily solved through effective iterative methods.

4. EXPERIMENTAL RESULTS

In this section, we first introduce the hyperspectral dataset used for the experiments. Then we report the performance of the proposed method and compare it against several baseline feature selection methods. The evaluation is based on classification results using support vector machine classifiers (SVM) [14].

Our experiments were performed on an APHI (Airborne Push hyperspectral Imager) image, which is shown in Fig 4(a). The ground truth is shown in Fig 4(b). This image is 210*150*64 in size, which covers 455nm to 805nm wavelength range. Some land-cover classes are discarded because of there are only very few training samples. The remaining surface covers are paddy, bamboo, tea, pachyrhizus, caraway and water, which are labeled as C4, T6, T7, V2, V13 and W2 in Fig. 4(b), respectively.

Fig1.a

Fig1.b

Fig. 1. (a) Band 60 of the hyperspectral image. (b) Ground truth of the land cover classes.

In order to evaluate the effectiveness of the proposed method, we compared it against two alternative band selection methods which include Maximum-variance PCA (MVPCA) [15] and Affinity propagation (AP) [5]. The MVPCA method ranks the bands based on the importance of individual band and its correlation with other bands. AP method groups bands based on their affinity, and select typical ones from groups. We also compare our method (HS) with the option of using all bands for classification (ALL).

In the process of semi-supervised learning, we have randomly selected 126 labeled samples and 2610 unlabeled samples as the training data. Numbers of labeled and unlabeled samples for each class are shown in Table 1.

Table 1. Number of labeled and unlabeled samples in the APHI data.

Class	Labeled samples	Unlabeled samples
Paddy	40	800
Bamboo	14	280
Tea	14	280
Pachyrhizus	10	250
Caraway	8	200
Water	40	800
Total	126	2610

After band selection, to test the accuracy of the methods being compared, an SVM classifier with RBF kernel was used for classification. We randomly selected 700 samples as the training data and 1600 samples as the testing data from the ground truth dataset, as shown in Table 2. The performance of each band selection method was evaluated by the classification accuracy. The experiments were run for ten times, with the averaged overall accuracy being reported to show the performance of the band selection methods.

Table 2. Numbers of training and testing samples for each class.

Class	Training samples	Testing samples
Paddy	200	400
Bamboo	100	200
Tea	50	150
Pachyrhizus	100	250
Caraway	50	200
Water	200	400
Total	700	1600

Fig. 2 shows the comparison results of our method against MVPCA, AP and ALL band selection methods. From the figure, we can observe that classification accuracies increase rapidly when the of number of bands are small, but slow down after 20 bands have been selected. In fact the results will converge to that of ALL method if all 64 bands are used. The figure also that the proposed HS method has clearly outperformed the MVPCA and the AP methods on the overall accuracy. This suggests that our method is very effective in selecting discriminative bands for classification.

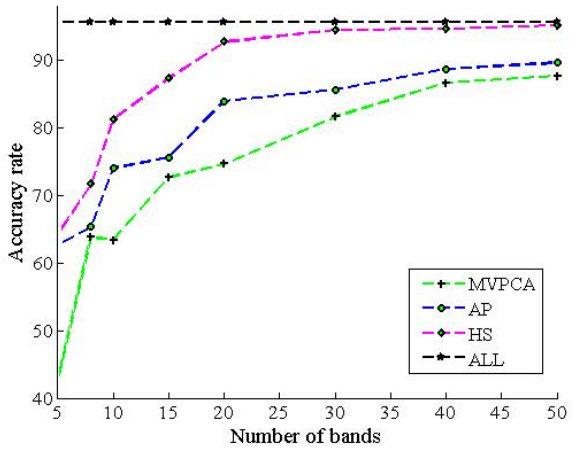

Fig. 2. Averaged overall accuracies by methods under comparison.

5. CONCLUSION

To conclude, we have presented a band selection model based on a sparse linear regression and a hypergraph models. It selects discriminative bands by ranking the sum of coefficients from the linear regression model. In order to address the problem of insufficient labeled data, hypergraph is used to compute an adjacency matrix, which is used to propagate labels to unlabeled data. The experimental results on the APHI data show that the proposed method has outperformed several baseline methods.

6. ACKNOWLEDGMENT

This research is supported under the National Natural Science Foundation of China projects No.61105002 and the Australian Research Councils DECRA Projects funding scheme (project ID DE120102948).

7. REFERENCES

[1] John Richards and Xiuping Jia, *Remote Sensing Digital Image Analysis*, Springer, 4th edition edition, 2006.

[2] Peter Bajcsy and Peter Groves, "Methodology for hyperspectral band selection," *Photogrammetric Engineering and Remote Sensing*, vol. 70, pp. 793–802, 2004.

[3] C.-I. Chang and S. Wang, "Constrained band selection for hyperspectral imagery," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 44, no. 6, pp. 1575–1585, 2006.

[4] B. Guo, S.R. Gunn, RI Damper, and JDB Nelson, "Band selection for hyperspectral image classification using mutual information," *Geoscience and Remote Sensing Letters, IEEE*, vol. 3, no. 4, pp. 522–526, 2006.

[5] Y. Qian, F. Yao, and S. Jia, "Band selection for hyperspectral imagery using affinity propagation," *IET Computer Vision*, vol. 3, no. 4, pp. 213–222, 2009.

[6] Qian Du, José M. Bioucas-Dias, and Antonio Plaza, "Hyperspectral band selection using a collaborative sparse model," in *Proceedings of the IEEE International Geoscience and Remote Sensing Symposium*, 2012, pp. 3062 – 3065.

[7] Ling Chen, Rui Huang, and Wei Huang, "Graph-based semi-supervised weighted band selection for classification of hyperspectral data," in *Audio Language and Image Processing (ICALIP), 2010 International Conference on*, 2010, pp. 1123–1126.

[8] Hongjun Su, Yehua Sheng, Peijun Du, and Kui Liu, "Adaptive affinity propagation with spectral angle mapper for semi-supervised hyperspectral band selection," *Applied Optics*, vol. 51, no. 14, pp. 2656–2663, 2012.

[9] R. Tibshirani, "Regression shrinkage and selection via the lasso," *Jounal of the Royal Statistical Society*, vol. 58, pp. 267–288, 1996.

[10] Yuntao Qian, Minchao Ye, and Jun Zhou, "Hyperspectral image classification based on structured sparse logistic regression and three-dimensional wavelet texture features," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 51, no. 4, pp. 2276–2291, 2013.

[11] C. Berge, *Hypergraphs Combinatorics of Finite Sets*, North-Holland, 1989.

[12] T. Sun Han, "Multiple mutual informations and multiple interactions in frequency data," *Information and Control*, vol. 46, no. 1, pp. 26–45, 1980.

[13] X. Zhu, Z. Ghahramani, J. Lafferty, et al., "Semi-supervised learning using gaussian fields and harmonic functions," in *Machine Learning-International Workshop Then conference*, 2003, vol. 20, p. 912.

[14] M. Pal and G.M. Foody, "Feature selection for classification of hyperspectral data by svm," *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 48, no. 5, pp. 2297–2307, 2010.

[15] C.I. Chang, Q. Du, T.L. Sun, and M.L.G. Althouse, "A joint band prioritization and band-decorrelation approach to band selection for hyperspectral image classification," *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 37, no. 6, pp. 2631–2641, 1999.