DISCRIMINATIVE WEIGHTED BAND SELECTION VIA ONE-CLASS SVM FOR HYPERSPPECTRAL IMAGERY

Yu Tang, Enlong Fan, Cheng Yan, Xiao Bai

School of Computer Science and Engineering
Beihang University
Haidian District, Beijing, China

Jun Zhou

Information and Communication Technology
Griffith University
Nathan, QLD, Australia

ABSTRACT

In the task of hyperspectral image classification, band selection is often adopted to select a band subset that is informative, thus reducing the computation and storage cost. To this end, we propose a supervised band selection method, which allows us to calculate the discriminative weight for each band. Specifically, we consider discriminative bands as bands that contribute more positive score to the classifier of the corresponding class than those of other classes when training. Based on this observation, we learn discriminative band weight vector for each class. Bands with larger discriminative weights are then selected for storage and classification. Our method can be efficiently solved in one-class SVM framework. The experimental results demonstrate the effectiveness of our method.

Index Terms— Band selection, hyperspectral imagery, image classification, one-class SVM, supervised learning

1. INTRODUCTION

Hyperspectral imaging sensors collect optical information from across the electromagnetic spectrum, usually generating hundreds of contiguous and narrow spaced spectral bands. Due to the phenomenon that certain objects have unique spectral signatures, hyperspectral imaging has great potential in a wide array of applications, such as mineral recognition, object detection [1] and land cover classification.

One challenging problem in hyperspectral imaging classification is that it is time-consuming and also requires expensive storage to process all the bands. Furthermore, high correlation exists among spectral bands, which implies that redundant bands could be removed without affecting classification performance. As a result, band dimensionality reduction technics are often used to address this problem, as a preprocessing step for hyperspectral images. Specifically, there are two kinds of such technics. The first one is feature extraction, which transforms the original space of raw hyperspectral images to a feature space with lower dimensions. Algorithms like PCA [2] and ICA [3] are suitable for this kind of

method. The drawback of feature extraction is that the transformed feature space no longer contains spectral information and lose the interpretability of hyperspectral data. The second one is feature selection or band selection, which chooses 'good' bands from the original ones. Here, 'good' means the bands are not redundant for classification. Various band selection methods have been proposed and they could be further separated into supervised [4], unsupervised [5] [6] and semi-supervised [7] manner. In applications that require high classification accuracy, supervised method is usually employed since it achieves better accuracy than unsupervised one.

Existing methods have different criteria for selecting bands. However, to achieve the best accuracy, the classification performance of each band should be considered when designing the criteria, which inspired us to optimize band weighting connected with classification results. Moreover, fully considering the discrimination among classes should further boost the performance.

In this paper, we propose a supervised band selection method based on max-margin optimization. We explicitly define discriminative bands of one class as bands that contribute more positive score to the classifier's output of this class than the classifiers' of other classes when training, which greatly influence the classification performance. Intuitively, these bands help make the correct classification. To formalize this observation, we use modified one-class SVM to satisfy our constraint and calculate discriminative band weights for each class. Bands with large weights are then selected. Our contributions are two-fold: First, we formalize the definition of discriminative band using the framework of one-class SVM classifier and the resulting objective function could be efficiently solved in standard SVM optimization tools. Second, our method calculate the discriminative weight for each band, which gives guidance to band selection.

2. APPROACH

In this section, we first introduce one-class SVM model, which will be used to learn discriminative weights for all bands. Then we describe how to design the constraint to

maintain the classification results while reducing the number of bands. Finally, we select bands with large discriminative weights and use them to train the classifier.

2.1. One-class SVM

To learn discriminative weights for hyperspectral bands, we adopt one-class SVM model. The classical SVM model is treated in a binary classification setting. For example, class label $y^i \in \{-1, 1\}$, which denotes class A and class B respectively. However, one-class SVM deals with anomaly detection problem, where only positive class training data could be used. The objective function of one-class SVM could be written as:

$$\min_{w_o} \sum_i \max(0, 1 - w_o^T x^i) + \alpha \|w_o\|_1 \quad (1)$$

Where w_o is the classification weight vector and α is the regularization parameter. The model learns w_o for positive class to discriminate positive class from non-positive class. Note (1) is different from the classical SVM in that we ignore class label y^i in the formula, since there is only positive class in one-class SVM model. In our method, we utilize the one-class SVM model to learn discriminative band weights by modifying the classification weights w_o .

2.2. Discriminative Band Selection

Let (x^i, y^i) be the i th sample in the training set, where $x^i \in \mathbb{R}^N$ is the vector of i th sample's intensity value of all bands and N is the total number of unselected bands, $y^i \in \{1, 2, \dots, C\}$ is the class label of the i th pixel and C is the number of classes. First, we train an initial classifier for each class in training set. In this step, we choose SVM but note that other classifiers such as multi-class logistic regression could be used as well. After the initial training, we have classification weights w_c for class c .

To get the correct classification result, the training sample should get higher score in the classifier of the correct class than those of other classes. For example, to be correctly classified, the i th sample must satisfy:

$$w_{y^i}^T x^i > w_c^T x^i, c \neq y^i \quad (2)$$

This gives us the insight of how to select discriminative bands. Intuitively, a band b of certain class is discriminative when the product of this band's intensity value and the classification weight of this certain class is higher than the product with any other classes, hence satisfying:

$$w_{y^i}(b)x^i(b) > w_c(b)x^i(b), c \neq y^i \quad (3)$$

which translates to

$$(w_{y^i}(b) - w_c(b))x^i(b) > 0, c \neq y^i \quad (4)$$

Our goal is to learn a N dimensional discriminative weight vector v for each class. Let $D^{c'}$ denotes the index of training subset with corresponding class label c' . We solve the following optimization problem for each class:

$$\min_v \sum_{i \in D^{c'}} \sum_{c \neq c'} \max(0, 1 - v^T u_{c',c}^i) + \lambda \|v\|_1 \quad (5)$$

where $u_{c',c}^i$ represents the difference value according to (4)

$$u_{c',c}^i(b) = (w_{c'}(b) - w_c(b))x^i(b) \quad (6)$$

We want to find discriminative bands, which satisfy the constraint of (4). Intuitively, (5) tries to estimate vector $v_{c'}$ such that the b th element of $v_{c'}$ represents the discriminative capability of band b to the c' class. In other words, band b of the c' class is discriminative to other classes if the b th element of $v_{c'}$ is large. In (5), the first term is a hinge loss term which measures discriminativeness of weight vector $v_{c'}$. For example, when $v_{c'}$ gives big weights to the discriminative bands, the loss term will be low in (5). The second term is a regularization term, which prevents the optimization from overfitting. Furthermore, we choose L1 norm as regularization so that sparsity is given to $v_{c'}$, which promotes the ability of band selection. The parameter λ controls the relative importance between hinge loss term and L1 regularization term. Note that (5) is equivalent to one-class SVM and could be efficiently solved by standard SVM solver [8].

After $v_{c'}$ is solved for each class, we select the n bands with the n largest weights. The selected bands contain the most discriminative information and are useful for classification.

The above steps are summarized in Algorithm 1.

Table 1. THE PSEUDO-CODE OF THE ALGORITHM

Algorithm 1

Input: Training set (x^i, y^i) , band number N , class number C

Output: Selected bands and corresponding classifiers

for Each class $c' = 1 \dots C$ do

 Train the initial classifier using N bands

 Use (5) to learn band weight vector $v_{c'}$

 Select bands according to $v_{c'}$

 Train classifier using selected bands

end for

3. EXPERIMENT

We evaluate our method on two publicly available datasets, Pavia Centre and Indian Pines. We report the experimental result of our method and compare it against other popular

technics. In all experiments we use linear SVM as both initial and final classifier, the regularization parameter in SVM and λ are obtained by cross-validation.

Pavia Centre is a section of the scene of Pavia, northern Italy, which was acquired by the ROSIS sensor during a flight campaign in 2002. It is a 1096x1096 pixels image and each pixel has 102 spectral bands. In Pavia Centre, we discard some samples since they contain no information. We randomly choose 10 percent of the remaining samples for each class as training set and the rest as test set. See Table 2 for detail.

Table 2. NUMBER OF TRAINING AND TESTING SAMPLES FOR EACH OBJECT CLASS IN PAVIA CENTRE

#	Class	Training	Testing
1	Water	83	741
2	Trees	82	738
3	Asphalt	82	734
4	Self-Blocking Bricks	81	727
5	Bitumen	81	727
6	Tiles	126	1134
7	Shadows	48	428
8	Meadows	83	741
9	Bare Soil	82	738
	TOTAL	748	6708

Fig. 1. Pavia Centre: (a) Sample band of the hyperspectral image. (b) Ground truth of the land cover class.

The second dataset we used is Indian Pines. Indian Pines is a section of the scene acquired by AVIRIS sensor in 1992 over the Indian Pines test site, North-western Indiana. It is a 145x145 pixels image and each pixel has 224 spectral bands covering wavelength ranging from 400nm to 2500nm. We remove water absorption bands: [104-108], [150-163], 220. Furthermore, we discard seven classes since only few samples are available for them, leaving nine classes available in the dataset. We randomly choose 10 percent of the samples for each class as training set and the rest as test set. See Table 3 for detail.

Fig. 2. SVM classification accuracy on Indian Pines by different methods.

Table 3. NUMBER OF TRAINING AND TESTING SAMPLES FOR EACH OBJECT CLASS IN INDIAN PINES

#	Class	Training	Testing
1	Corn-no till	143	1285
2	Corn-mil till	83	747
3	Grass/pasture	48	435
4	Grass/trees	73	657
5	Hay-windrowed	48	430
6	Soybean-no till	97	875
7	Soybean-mil till	246	2209
8	Soybean-clean till	59	534
9	Woods	127	1138
	TOTAL	924	7172

We also compare our method with two popular band selection technics AP [9], MVPCA [10]. AP uses affinity propagation to calculate the similarity between bands and search for exemplars as selected bands. MVPCA selects bands by ranking them with the importance of each individual band and its correlation with others. Classification with full bands is reported as well. Fig.2 and Fig.4 show the results on the two datasets respectively. When the number of selected bands increases, the classification accuracy also increases and gets close to the full bands after certain number been reached. It can be observed that our method outperforms the alternatives in both datasets, which demonstrates the advantage of our method in classification accuracy.

4. CONCLUSION

In this paper, we proposed a new method for hyperspectral band selection. The success of our method is due to the following reasons: 1) discriminative band is defined explicitly

Fig. 3. Indian Pines: (a) Sample band of the hyperspectral image. (b) Ground truth of the land cover class.

Fig. 4. SVM classification accuracy on Indian Pines by different methods.

which fully taking the discrimination among classes into account, 2) we use modified one-class SVM to formulize the optimization problem so that discriminative weight vector for each class can be efficiently solved. The experimental results on two widely used datasets, Pavia Centre and Indian Pines, show the effectiveness of our method.

5. ACKNOWLEDGEMENT

This work was supported by NSFC project No. 61370123 and BNSF project No. 61503422, and the Australian Research Councils DECRA Projects funding scheme (project ID 522 DE120102948).

6. REFERENCES

- [1] Xiao Bai, Huigang Zhang, and Jun Zhou, “Vhr object detection based on structural feature extraction and query expansion,” *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 52, no. 10, pp. 6508–6520, 2014.
- [2] Abhishek Agarwal, Tarek El-Ghazawi, Hesham El-Askary, and Jacqueline Le-Moigne, “Efficient hierarchical-pca dimension reduction for hyperspectral imagery,” in *Signal Processing and Information Technology, 2007 IEEE International Symposium on*. IEEE, 2007, pp. 353–356.
- [3] Qian Du and He Yang, “Similarity-based unsupervised band selection for hyperspectral image analysis,” *Geoscience and Remote Sensing Letters, IEEE*, vol. 5, no. 4, pp. 564–568, 2008.
- [4] He Yang, Qian Du, Hongjun Su, and Yehua Sheng, “An efficient method for supervised hyperspectral band selection,” *Geoscience and Remote Sensing Letters, IEEE*, vol. 8, no. 1, pp. 138–142, 2011.
- [5] Baofeng Guo, Steve 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.
- [6] Xiurui Geng, Kang Sun, Luyan Ji, and Yongchao Zhao, “A fast volume-gradient-based band selection method for hyperspectral image,” *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 52, no. 11, pp. 7111–7119, 2014.
- [7] Xiao Bai, Zhouxiao Guo, Yanyang Wang, Zhihong Zhang, and Jun Zhou, “Semisupervised hyperspectral band selection via spectral-spatial hypergraph model,” *Selected Topics in Applied Earth Observations and Remote Sensing, IEEE Journal of*, vol. 8, no. 6, pp. 2774–2783, 2015.
- [8] Chih-Chung Chang and Chih-Jen Lin, “Libsvm: A library for support vector machines,” *ACM Transactions on Intelligent Systems and Technology (TIST)*, vol. 2, no. 3, pp. 27, 2011.
- [9] Yuntao Qian, Fengqi Yao, and Shenli Jia, “Band selection for hyperspectral imagery using affinity propagation,” *IET Computer Vision*, vol. 3, no. 4, pp. 213–222, 2009.
- [10] Chein-I Chang, Qian Du, Tzu-Lung Sun, and Mark 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.

[1] Xiao Bai, Huigang Zhang, and Jun Zhou, “Vhr object detection based on structural feature extraction and query expansion,” *Geoscience and Remote Sensing,*