Classification Oriented Semi-supervised Band Selection for Hyperspectral Images

Jun Bai, Shiming Xiang and Chunhong Pan
Institute of Automation, Chinese Academy of Sciences, Beijing
e-mail: {jbai, smxiang, chpan}@nlpr.ia.ac.cn

Abstract

This paper proposes a new framework of band selection for object classification in hyperspectral images. Different from traditional approaches where the selected bands are shared from all classes, in this work, different subsets of bands are selected for different class pairs. Without prior knowledge of spectral database, we estimate the spectral characteristic of objects with the labeled and unlabeled samples, benefiting from the concept of semi-supervised learning. Under the assumption of Gaussian mixture model (GMM), the vectors of mean values and covariance matrices for each class are estimated. The separabilities for all pairs of classes are thus calculated on each band. The bands with the highest separabilities are then selected. To validate our band selection result, support vector machine (SVM) is employed using a strategy of one against one (OAO). Experiments are conducted on a real data set of hyperspectral image, and the results can validate our algorithm.

1. Introduction

In hyperspectral image processing, band selection is a very hot topic that has been studied for a decade. However, the motivation behind the scene may be quite various. Due to the limitations of hardware, early studies aim to reduce the burdens of calculations. Among these uniform band selection is an easy but typical method [2]. In this method, the set of spectral bands is uniformly sampled to a subset with pre-defined number of bands. As a sacrifice, the loss of information for this algorithm is usually very large. Some algorithms calculate the correlations between the spectral bands and reserve the bands with least correlations or reject some highly correlated bands to reduce the redundancy [10], [11]. Among these algorithms an energy function is usually established to describe the correlation [3]. Accompanied with these band selection methods, a band clustering algorithm is usually performed as well [8].

Another family of algorithms is more closely designed for the purpose of classification. These algorithms aim to achieve the highest separabilities between classes on the select subset of bands [9], [5], [7]. To reach this goal, labeled samples are utilized in supervised band selection, or more recently, both labeled and unlabeled samples are used in semi-supervised band selection. The separabilities are defined by divergence or a certain form of distance, e.g, Bhattacharyya distance, Jeffries-Matusita (JM) distance. Unfortunately, these separabilities based algorithms select the same subset of bands for all class combinations and ignore the possibility that is easy to think of: for different pairs of classes, can the subsets of bands that best describe the separabilities be different?

In this paper, we design the band selection algorithm specially to the purpose of classification, and propose an entirely new band selection strategy. Different from the state-of-art algorithms, the selected subsets of bands can be different for different pairs of classes. To achieve this goal, we select spectral bands for all pairs of classes separately. For each class pair, the separabilities are evaluated according to the difference of spectral characteristics on each band. To reach this goal, it is necessary to obtain spectral characteristics for each class.

For a long time, an ideal assumption is that a spectral database is available. By contrasting with the database, the detailed information for each class of objects can be acquired, as well as spectral characteristics. In this paper, however, we utilize the labeled samples, and some unlabeled samples. This is so called a semi-supervised band selection method. With these samples, the statistical parameters for each class are estimated. These parameters are the evidence for the separabilities between the classes.

The remainder of this paper is organized as follows. Section 2 describes the framework of our work. In sec-
tion 3 the details of our methodology are presented. The experimental results are demonstrated in section 4. The conclusions are drawn in the final section.

2. The main framework

As is described in section 1, our band selection algorithm is designed for a classification task. For this reason, the core idea is to increase the separabilities between the classes. In fact, the separabilities between a pair of classes often concentrate on a subset of bands, and these subsets may be quite different for different pair of classes. This idea is what our algorithm originally arises from.

Our problem can be described as: given a data set of hyperspectral image, and a set of labeled and unlabeled samples \( \mathcal{X}^+ \) and \( \mathcal{X}^- \), the goal of our formulations is to find a collection \( S = \{ S_{ij} : i, j \in L, i \neq j \} \), where \( S_{ij} \) is a subset of bands and \( L \) is the label set.

To accomplish our work, we establish our algorithm on the basic assumption that the distributions of the spectral data fit a solid model, or GMM. From this model, we can easily estimate some statistical parameters, such as mean values and covariance matrices.

The main steps of our algorithm can be expressed as follows.

1. Benefiting from the algorithm of Semi-supervised Expectation Maximization (SS-EM) algorithm [1], we use labeled and some unlabeled samples to estimate the mean values and covariances for all classes on all the bands.

2. For each pair of classes, we estimate the separabilities on each band with the calculated statistical parameters.

3. For each pair of classes, all the bands are sorted by their separabilities. To achieve the best classification result, we select the top \( n \) bands with the highest separabilities, where \( n \) is a pre-defined number for real applications.

In this work, we use spectral features only and no spatial information is considered.

3. Methodologies and validations

3.1. SS-EM

Although expectation maximization (EM) has been proposed for years, the research and applications of SS-EM have not gained attention until the recent years. In this paper, we adopt the SS-EM algorithm proposed in [1]. Meanwhile, we make some improvements to reject the outliers.

Compared with traditional unsupervised EM, in SS-EM some samples are assigned solid labels. The optimization of log-likelihood function thus turns out to be:

\[
\log L(\mathcal{X}, \mathcal{Z}; \Psi) = \sum_{j \in \mathcal{X}^+} \sum_{i=1}^{c} z_{ji} \log \pi_i f_i(x_j; \theta_i) + \sum_{j \in \mathcal{X}^-} \log \pi_{l(j)} f_{l(j)}(x_j; \theta_{l(j)})
\]

In (1) \( \mathcal{X} \) is the set of samples, and \( \mathcal{X}^+ \) and \( \mathcal{X}^- \) are the sets of labeled and unlabeled samples separately, \( \pi_i \) is the prior probability, \( z_{ji} = 1 \) if \( x_j \) is a sample of class \( i \) and 0 otherwise. \( f_i(x_j; \theta_i) \) is the probability function and \( \theta_i = (\mu_i, \Sigma_i) \) are mean values and covariance matrices to be estimated, and \( l(j) \) stands for the label of sample \( x_j \).

To clarify the optimization process of SS-EM, we review the process of traditional EM as E-step (2) and M-step (3)-(5) separately:

\[
\tau_{ji}^{(k+1)} = \frac{\pi_i^{(k)} f_i(x_j; \hat{\theta}_i^{(k)})}{\sum_{h=1}^{N} \pi_h^{(k)} f_h(x_j; \hat{\theta}_h^{(k)})} \quad (2)
\]

\[
\pi_i^{(k+1)} = \frac{\sum_{j=1}^{N} \tau_{ji}^{(k+1)}}{N} \quad (3)
\]

\[
\hat{\mu}_i^{(k+1)} = \frac{\sum_{j=1}^{N} x_j \tau_{ji}^{(k+1)}}{\sum_{j=1}^{N} \tau_{ji}^{(k+1)}} \quad (4)
\]

\[
\Sigma_i^{(k+1)} = \frac{\sum_{j=1}^{N} (x_j - \hat{\mu}_i^{(k+1)}) (x_j - \hat{\mu}_i^{(k+1)})^T \tau_{ji}^{(k+1)}}{\sum_{j=1}^{N} \tau_{ji}^{(k+1)}} \quad (5)
\]

In (3) \( \tau_{ji} \) is an estimation of \( z_{ji} \) or the posterior probability, and \( k \) is the number of iterations.

In [1], it is pointed out that the main difference of SS-EM with respect to traditional EM is the estimation of posterior probability for the labeled samples. For a labeled sample, it is not formulated as (2). Instead, it is defined as:

\[
\tau_{ji} = \begin{cases} 1, & \text{if } i = l(j) \\ 0, & \text{otherwise} \end{cases}
\]

Another problem is the initialization. In this semi-supervised context, the parameters can be initialized by the labeled samples.
3.2. Separability estimation and band selection

Under the assumption of GMM, the samples for each class obey the distribution of \( d \)-dimensional Gaussian. As a mathematical rule, the marginal distribution on each band is also 1-dimensional Gaussian. Having the statistical information for each class, we now concentrate our problem on a certain pair of classes. On each band, the next task to be done is to estimate the 1-dimensional separability between the samples of two classes. In fact, this separability can be expressed as the Bhattacharyya distance as a classical pattern recognition problem [6]. Let \((\mu_1, \sigma_1^2)\) and \((\mu_2, \sigma_2^2)\) be the means and covariances separately for the two classes, the separability can be estimated as:

\[
D = \frac{1}{4} \left( \mu_1 - \mu_2 \right)^2 + \frac{1}{2} \log \left( \frac{\sigma_1^2 + \sigma_2^2}{2\sigma_1 \sigma_2} \right) \quad (7)
\]

The remaining problem is how to select the subset of bands for the class pair. In this paper we use a simple strategy: by sorting all the bands with their separabilities, we select those bands with the highest separabilities. The number of bands can be predefined to meet the requirement of the application.

3.3. Validations

Due to the fact that our band selection algorithm is designed for classification, the results of our work are also validated by classification tasks. Since the selected bands are mostly different for all pairs of classes, the decision of a testing sample is made with a binary classifier on any possible combinations of class pairs with the corresponding selected bands. Then a voting rule is performed via the well known one against one strategy (OAO). In theory, any binary classifier can accomplish this task. In this work, we use the SVM classifier and the software is LibSVM [4].

4. Experimental results

The data set, Pavia city is provided by the HySens project acquired by the ROSIS-3 optical sensor. For our experiment, we cut off the part on the left of the image and retain the part on the right since there is a black strip for unknown regions. The number of bands is 102 with a spectral range from 0.43\(\mu m\) to 0.86\(\mu m\). The spatial resolution is about 1.3\(m\) per pixel. Its size is 1096 \(	imes\) 492. Nine classes are labeled: water, asphalt, trees, shadow, meadows, bare soil, tiles, bricks and bitumen. From all the bands we select three to combine a false RGB image in Figure 1(a). The training (labeled) and testing samples for classification are also presented in Figure 1 (b) and (c).

According to the algorithm described in section 3, we perform our band selection process. An SVM classifier is employed to validate the results of band selection. To evaluate the performance of our algorithm, we test with different values for the number of bands \(n\), e.g., 10, 20, 30, ..., 90. The change of overall accuracies with the number of selected bands are illustrated in Figure 2.

In Figure 2, when the number of selected band is small \((n = 10)\), the result is already fine. This shows very high separabilities in this case. The overall accuracy of the classification results increases with the number of bands. From this figure, when the number of band is 60, the overall accuracy is higher than that attained with all bands. This proves the fact that the separabilities can concentrate on the subsets of bands. This is also because there are some noisy bands. From these analysis this figure can well validate our algorithm.

Specially we test the result when the number of selected band is small on an extreme case. That is when \(n = 3\). The classification result with all bands is also testified. For comparisons we also show the classification results of the uniform method for \(n = 3\), which are reported in Table 1.

In Table 1, it is illustrated that the classification results can exhibit rather well. The results for some classes can approach those with all the bands. Compared to the uniform method, the results for some classes behave much better. This is because we achieve better separabilities using our band selection algorithm.
Table 1. Classification accuracies of the band selection results for the AVIRIS data set when $n = 3$

<table>
<thead>
<tr>
<th>Classes</th>
<th>Uniform</th>
<th>Ours</th>
<th>All Bands</th>
</tr>
</thead>
<tbody>
<tr>
<td>Water</td>
<td>97.06</td>
<td>96.88</td>
<td>96.74</td>
</tr>
<tr>
<td>Asphalt</td>
<td>93.13</td>
<td>93.79</td>
<td>95.79</td>
</tr>
<tr>
<td>Trees</td>
<td>78.72</td>
<td>86.23</td>
<td>92.47</td>
</tr>
<tr>
<td>Shadow</td>
<td>99.63</td>
<td>98.66</td>
<td>99.95</td>
</tr>
<tr>
<td>Meadows</td>
<td>93.59</td>
<td>93.66</td>
<td>96.34</td>
</tr>
<tr>
<td>Bare soil</td>
<td>96.95</td>
<td>93.71</td>
<td>96.00</td>
</tr>
<tr>
<td>Tiles</td>
<td>29.79</td>
<td>61.11</td>
<td>98.99</td>
</tr>
<tr>
<td>Bricks</td>
<td>58.69</td>
<td>50.05</td>
<td>83.74</td>
</tr>
<tr>
<td>Bitumen</td>
<td>74.13</td>
<td>78.81</td>
<td>95.70</td>
</tr>
<tr>
<td>OA</td>
<td>91.04</td>
<td>92.37</td>
<td>96.14</td>
</tr>
</tbody>
</table>

Figure 2. The variations of overall accuracy with the number of selected bands

5. Conclusion

This paper presents a new framework of band selection, which we call classification oriented band selection. Different from traditional band selection algorithm, the subsets of bands are selected customized to the class pairs. To estimate the separabilities between band, a semi-EM algorithm is employed, and the statistical parameters are estimated. Finally the classification results are demonstrated. For comparison, the results of the uniform method are also presented. These results can validate our method.

6. Acknowledgment

This work is supported by the National Basic Research Program of China (Grant No. 2012CB316304), and the Sustainable Urban Metabolism for Europe project (SUME, Grant Agreement No: 212034). The authors would like to thank the University of Pavia and the HySenS project for providing the data.

References