International Journal of Latest Engineering Research and Applications (IJLERA) ISSN: 2455-7137

Volume –06, Issue – 09, September 2021, PP – 16-23

Maximum Margin Criterion and Manifold Learning based Hyperspectral Image Classification

Yanbin Peng[1], Zhijun Zheng[1]

[1]*School of Information and Electronic Engineering, Zhejiang University of Science and Technology, Hangzhou 310023, China*

Abstract: Hyperspectral image classification has drawn more and more attention recently. However, the high number of spectral bands and the low number of labeled sample points hamper the increasing of classification precision of the hyperspectral images. This paper presents a maximum margin criterion and manifold Learning (MCML) based hyperspectral image classification method. MCML method has two advantages: (1) it is a supervised dimensionality reduction method, which improves the discriminative ability of new feature set by exploiting label information of original training sample points; (2) it is a nonlinear method that projects the original feature set into a new feature set by preserving the local relation between sample points. The proposed method is applied to hyperspectral image classification and is examined using real hyperspectral data sets. Experimental results show that it is more suitable for hyperspectral image classification tasks than the other methods.

Keywords: Hyperspectral image; classification; maximum margin criterion; manifold learning

1. Introduction

Currently, hyperspectral image classification has attracted considerable interest in pattern recognition of remote sensing fields [1-3]. It is well known that the dimension of pixels in the hyperspectral images is usually very high. For instance, the feature of a pixel in AVIRIS Indian pine data is a 220-dimensional vector. High dimensionality of feature vector has led to a series of problems. Firstly, the higher number of features the pixel has, the more complex the classification algorithm becomes, which leads to the increment of computational complexity. Secondly, the bands in the hyperspectral images are correlated. The redundant feature may degrade the performance of the classifier. At last the labeling cost is too high, which causes the scarce of labeled training sample points. In a word, the "curse of dimensionality" problem hampers the improvement of classification efficiency in hyperspectral image recognition. The contradiction of high dimensional and low training samples is an open question even in machine learning [4, 5]. Dimensionality reduction algorithm is a common method to solve this problem. Nowadays, an enormous volume of literature has been devoted to investigating dimensionality reduction methods. The traditional dimensionality reduction methods can be classified into two categories: feature selection and feature abstraction.

Feature selection methods select feature subset from the original feature set through a kind of criterion. Scholars had put forward their criteria. Peng [6] selects hyperspectral bands through the spectral clustering method. Tan [7] proposes a simple unsupervised framework to effectively select and combine spectral information and spatial features for support vector machine based classification. Li [8] proposes novel group sparsity based semi-supervised band selection method. Guo[9] presents a semi-supervised learning approach and a hypergraph model to select useful bands based on few labeled object information.

Compared to the feature selection method, the feature abstraction method generally achieves higher accuracy in classifying. Feature abstraction tries to project the original feature space into a new and reduced feature space. Linear feature abstraction method includes principal component analysis (PCA for short) and linear discriminant analysis (LDA for short) [10, 11]. PCA conducts Eigen-decomposition of covariance matrix in the original data and then projects the original data into a subspace spanned by the eigenvectors with the largest eigenvalues. LDA is a supervised method, which incorporates label information of original data in the process of projection. Linear feature abstraction methods only consider linear global Euclidean structure.

However, recent research shows that the hyperspectral pixel may reside on a nonlinear low-dimensional manifold structure. Thus, the linear method is not suited for the classification of hyperspectral pixels. In a nonlinear field, the kernel-based algorithm can map the original nonlinear observations into a higher-dimensional kernel space, where the classification problem becomes linearly separable. Li [12] presents a new framework for developing generalized composite kernel machines for hyperspectral image classification. However, the success of pixel classification is depended on the choice of the kernel. In general, selecting an efficient kernel is up to the skills and experiences of the human users. Another nonlinear method is manifold learning. Manifold learning considers that there is a low-dimensional manifold that is embedded in original high-dimensional space. Therefore, the manifold learning method tries to find a different geometrical structure while preserving the locality proximity relationship among the original data points.

The existing methods suffer from a limitation that it does not comprehensively consider discriminant information and subspace projection, which is very important for hyperspectral image classification. In this paper, we propose a maximum margin criterion and manifold learning (MCML for short) based hyperspectral image classification method. MCML method has two advantages: (1) it is a supervised dimensionality reduction method, which improves the discriminative ability of new feature set by exploiting label information of original training sample points; (2) it is a nonlinear method that projects the original feature set into a new feature set by preserving the local relation between sample points. The proposed method is applied to hyperspectral image classification and is examined using real hyperspectral data sets. Experimental results show that it is more suitable for hyperspectral image classification tasks than the other methods.

The remainder of this paper is organized as follows: in section 2, we will describe the characteristic of hyperspectral image data. A detailed introduction of the MCML method is given in section 3. The experimental results for applying our method to hyperspectral image classification will be presented in section 4, followed by the conclusions in section 5.

2. Hyperspectral Image Cube

The Hyperspectral image can be expressed as a three-dimensional data cube. As is shown in figure 1, dimensions I and J corresponds to the width and length dimensions of hyperspectral data. Dimension K corresponds to the spectral dimension. There is K band images in this data, where d_k is the kth band image. Pixel is expressed as a K dimension vector p . Hyperspectral data set is expressed as: $\boldsymbol{P} = [\boldsymbol{p}_1, \boldsymbol{p}_2, ..., \boldsymbol{p}_n]$. $\boldsymbol{M}$ represents the projection matrix.

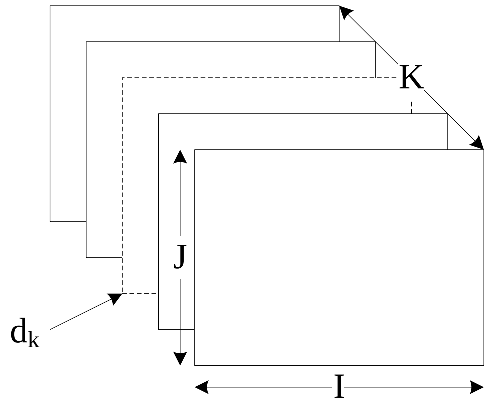

Figure 1. Hyperspectral image cube

3. Maximum margin criterion and manifold learning method

Maximum margin criterion [13] is a classic dimension reduction technique for supervised learning. MMC hopes that the data points of the same category are as close as possible in the projection space, and the centers of data points of different categories are as far apart as possible. The variance within a class is minimized and the variance between classes is maximized.

In MMC, the objective function can be expressed as:

$$\mathrm{O}=\max\{\sum_{ij} r_i r_j (\mathrm{k}(\mathrm{s_i},\mathrm{s}_j)-\mathrm{t}(\mathrm{s_i})-\mathrm{t}(\mathrm{s}_j))\} \tag{1}$$

Wherein r_i represents prior probability of class i. $\mathrm{s_i}$ is mean vector of class i.

$$\mathrm{k}(\mathrm{s_i},\mathrm{s}_j)=\|\mathrm{s_i}\text{-}\mathrm{s_j}\|,\ \mathrm{t}(\mathrm{s_i})=\mathrm{trace}(\mathrm{C}_i),\ \mathrm{t}(\mathrm{s_j})=\mathrm{trace}(\mathrm{C_j})$$

Wherein C_i is covariance matrix of class i.

Thus the optimized objective function can be derived as:

$$\mathrm{O}=\min(\boldsymbol{M}^{\mathrm{T}}\boldsymbol{Q}^{MMC}\boldsymbol{M}) \tag{2}$$

Wherein $\boldsymbol{Q}^{MMC}=(\boldsymbol{Q}^{w}\text{-}\boldsymbol{Q}^{b})$. $\boldsymbol{Q}^{w}$ is within-class scatter matrix. $\boldsymbol{Q}^{\mathrm{b}}$ is between-class scatter matrix.

MMC ensures the compactness of data within classes by minimizing variance within classes and the separability of data by maximizing variance between classes. MMC retains the discriminant information of training sample points when solving the projection matrix.

However, the calculation process is carried out in the Euclidean space. Recent studies show that the sample points in higher dimensional space are usually not distributed in Euclidean space but show nonlinear manifold structure. There are embedded low-dimensional manifold structures in high-dimensional space. Manifold learning theory assumes that the sample points are distributed on a low-dimensional fluid body embedded in a high-dimensional space. It is considered that sample points with similar distances in the original space are also similar in the projection space, that is, the local invariance hypothesis. The manifold learning method can recover low dimensional manifold structure from high dimensional sample points and give corresponding mapping function from high dimensional to low dimensional. Classical manifold learning methods include Isomap, Laplacian eigenmap (LE for short) and local linear embedding (LLE for short). In this paper, we use the manifold learning method to regularize the MMC objective function so that the projection matrix not only retains the discrimination information of training sample points but also retains the manifold geometry structure.

Manifold learning severely penalizes points with close distances in the original space for being far apart in the projection space. Construct the nearest neighbor graph $\boldsymbol{G}=(\boldsymbol{P},\boldsymbol{W}^{\mathrm{m}})$, wherein $\boldsymbol{P}$ is the set of sample points, $\boldsymbol{W}^{\mathrm{m}}$ is the weight matrix between sample points. The weight matrix is defined as follows:

$$\boldsymbol{W}_{ij}^{\mathrm{m}}=\begin{cases} w_{ij} & if\ \boldsymbol{p}_i \in \mathrm{KNN}(\boldsymbol{p}_j)\ or\ p_j \in \mathrm{KNN}(\boldsymbol{p}_i) \\ 0 & otherwise \end{cases} \tag{3}$$

Wherein $\mathrm{KNN}()$ represents the k nearest neighbor algorithm, w_{ij} represents the similarity between sample points $\boldsymbol{p}_i$ and $\boldsymbol{p}_j$. Due to the high dimensional characteristics of sample points, the similarity calculation method based on Euclidean distance is not suitable for calculating w_{ij}. This paper proposes a method based on divergence to calculate similarity. Assuming that $\mathrm{h}_i(b)$ and $\mathrm{h}_j(b)$ is the gray histogram of the sum of sample points $\boldsymbol{p}_i$ and $\boldsymbol{p}_j$, and b is the gray value, the KL divergence of $\boldsymbol{p}_i$ and $\boldsymbol{p}_j$ is defined as follows:

$$\mathrm{KLdivergence}(i,j)=\sum_{b=1}^{\mathrm{B}}(\mathrm{h}_i(b)\log\frac{\mathrm{h}_i(b)}{\mathrm{hist}_j(b)}) \tag{4}$$

Since the divergence value is not symmetric, the similarity of sample points $\boldsymbol{p}_i$ and $\boldsymbol{p}_j$ is defined as:

$$w_{ij}=(1-\frac{1}{2}(\sum_{b=1}^{\mathrm{B}}(\mathrm{h}_i(b)\log\frac{\mathrm{h}_i(b)}{\mathrm{h}_j(b)}) + \sum_{b=1}^{\mathrm{B}}(\mathrm{h}_j(b)\log\frac{\mathrm{h}_j(b)}{\mathrm{h}_i(b)}))) \tag{5}$$

The corresponding manifold regular term is defined as follows:

$$\arg\min_{\boldsymbol{M}}\sum_{ij}\|\boldsymbol{M}^{\mathrm{T}}\boldsymbol{p}_i-\boldsymbol{M}^{\mathrm{T}}\boldsymbol{p}_j\|^2\boldsymbol{W}_{ij}^{\mathrm{m}} \tag{6}$$

The derivation of the regular term objective function is as follows:

$$\begin{aligned}&\sum_{ij} \| \boldsymbol{M}^{\mathrm{T}} \boldsymbol{p}_i - \boldsymbol{M}^{\mathrm{T}} \boldsymbol{p}_j \|^2 \boldsymbol{W}_{ij}^{\mathrm{m}} \\ &= \boldsymbol{M}^{\mathrm{T}} \boldsymbol{P} (\boldsymbol{D}^{\mathrm{m}} - \boldsymbol{W}^{\mathrm{m}}) \boldsymbol{P}^{\mathrm{T}} \boldsymbol{M} \\ &= \boldsymbol{M}^{\mathrm{T}} \boldsymbol{P} \boldsymbol{L}^{\mathrm{m}} \boldsymbol{P}^{\mathrm{T}} \boldsymbol{M} \\ &= \boldsymbol{M}^{\mathrm{T}} \boldsymbol{Q}^{\mathrm{m}} \boldsymbol{M}\end{aligned}$$

Wherein $\boldsymbol{D}^{\mathrm{m}}$ is the diagonal matrix, $\boldsymbol{D}_{ii}^{\mathrm{m}} = \sum_j \boldsymbol{W}_{ij}^{\mathrm{m}}$, $\boldsymbol{L}^{\mathrm{m}} = \boldsymbol{D}^{\mathrm{m}} - \boldsymbol{W}^{\mathrm{m}}$ is the corresponding Laplace matrix, $\boldsymbol{Q}^{\mathrm{m}} = \boldsymbol{P}\boldsymbol{L}^{\mathrm{m}}\boldsymbol{P}^{\mathrm{T}}$.

Considering maximum margin criterion and manifold learning, the objective function of the dimension reduction algorithm (MCML) in this paper is defined as follows:

$$\begin{aligned}&\min_{\boldsymbol{M}} \boldsymbol{M}^{\mathrm{T}} \boldsymbol{Q} \boldsymbol{M} \\ &s.t.\ \boldsymbol{M}^{\mathrm{T}} \boldsymbol{P} \boldsymbol{P}^{\mathrm{T}} \boldsymbol{M} = \boldsymbol{I}\end{aligned} \tag{7}$$

Wherein $\boldsymbol{Q} = (\boldsymbol{Q}^{\mathrm{MMC}} + \boldsymbol{Q}^{\mathrm{m}})$

The optimization problem can be transformed into a generalized feature decomposition problem as follows:

$$\boldsymbol{Q}\boldsymbol{m} = \lambda \boldsymbol{D}\boldsymbol{D}^{\mathrm{T}} \boldsymbol{m} \tag{8}$$

The eigenvector corresponding to the minimum q eigenvalues is the solution of optimization problem (7), expressed as $\boldsymbol{M} = [\boldsymbol{m}_1, \boldsymbol{m}_1, ..., \boldsymbol{m}_q]$. Pseudo-codes based on maximum margin criterion and manifold learning algorithm (MCML) is as follows:

Algorithm 1: MCML

Input: Training sample set: $\boldsymbol{D} = [\boldsymbol{p}_1, \boldsymbol{p}_2, ..., \boldsymbol{p}_n]$

Output: projection matrix: $\boldsymbol{M} = [\boldsymbol{m}_1, \boldsymbol{m}_1, ..., \boldsymbol{m}_q]$

(1) Calculate the matrix, $\boldsymbol{Q}^{\mathrm{MMC}}$, $\boldsymbol{Q}^{\mathrm{m}}$, and $\boldsymbol{Q}$;

(2) Feature decomposition is carried out for the problem (8), and the eigenvectors $\boldsymbol{M}$ corresponding to the smallest q eigenvalues are formed.

4. Experiments

Having presented the method of maximum margin criterion and manifold learning based hyperspectral image classification method (MCML for short) in the previous sections, this paper demonstrates the effect of the proposed method through several comparative experiments. These experiments are done in a real-world hyperspectral image data set.

Figure 2. ROSIS Pavia data

The data was collected by the ROSIS sensor, which was centered at the University of Pavia. The spectral range is from 0.43 micrometer to 0.86 micrometers. There are totally 610×340 pixels in this image. It has 115 spectral bands, wherein 103 bands were used in the experiment. The other 12 bands were removed as noise bands. The pseudocolor scene of the data is shown in figure 2. There are nine classes of land cover. The type and number of training samples and test samples are detailed listed in table 1.

Table 1. Number of training and testing samples in the ROSIS Pavia data

Class		Number of Samples	
No	Name	Training	Testing
1	Asphalt	200	5000
2	Meadows	200	10000
3	Gravel	200	1500
4	Trees	200	2000
5	Metal sheets	200	1000
6	Bare soil	200	4000
7	Bitumen	200	1000
8	Bircks	200	3000
9	Shadows	200	800

In order to assess the performance of the MCML method, we choose three models for comparison: 1) SDE algorithm proposed in literature [14]; 2) LGGSP algorithm proposed in literature [15]; 3) SDLEA algorithm proposed in literature [16].

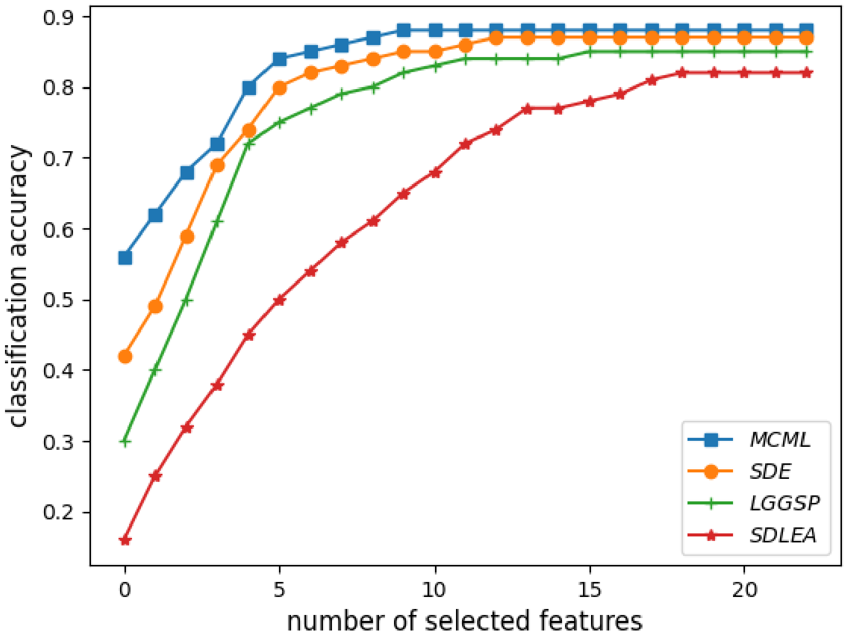

Figure 3. Comparison of classification accuracy in ROSIS Pavia data

Figure 3 displays the classification precision of algorithm MCML, SDE, LGGSP, and SDLEA, respectively. As is shown in figure 3, the MCML algorithm achieves the best performance.

5. Conclusions

In this paper, we investigate the combination of discriminant information and neighbor information for dimensionality reduction in hyperspectral image classification. In order to make use of label information, MMC is imported into the optimization objective function to maximize the margin between different classes. Then, the manifold regularization term is used to preserve the manifold geometry structure. Experimental results show that the proposed method is more suitable for hyperspectral image classification tasks than the other methods.

References:

[1]. Kalaiarasi G , Maheswari S . Deep proximal support vector machine classifiers for hyperspectral images classification[J]. Neural Computing and Applications, 2021(3).

[2]. Go A , Go B . Comparative study of hyperspectral image classification by multidimensional Convolutional Neural Network approaches to improve accuracy[J]. Expert Systems with Applications, 2021

[3]. Pathak D K , Kalita S K , Bhattacharya D K . Spectral spatial joint feature based convolution neural network for hyperspectral image classification[J]. Concurrency and Computation: Practice and Experience, 2021.

[4]. Nan L , Chee M L , Zhi X K , et al. Utilizing machine learning dimensionality reduction for risk stratification of chest pain patients in the emergency department[J]. BMC Medical Research Methodology, 2021, 21(1).

[5]. Kahnzadi P , Majidi B , Adabi S , et al. Robust fuzzy rough set based dimensionality reduction for big multimedia data hashing and unsupervised generative learning[J]. Multimedia Tools and Applications, 2021, 80(12):17745-17772.

[6]. Peng Y B, Jie qing AI. Hyperspectral Imagery Classification Based on Spectral Clustering Band Selection[J]. Opto-Electronic Engineering,2012,39(2):63-67.

[7]. Tan K, Li E, Du Q, et al. Hyperspectral Image Classification Using Band Selection and Morphological Profiles[J]. IEEE Journal of Selected Topics in Applied Earth Observations & Remote Sensing, 2014, 7(1):40-48.

[8]. Li H, Wang Y, Duan J, et al. Group sparsity based semi-supervised band selection for hyperspectral images[C]// Image Processing (ICIP), 2013 20th IEEE International Conference on. IEEE, 2013:3225-3229.

[9]. Guo Z, Bai X, Zhang Z, et al. A hypergraph based semi-supervised band selection method for hyperspectral image classification[C]// Image Processing (ICIP), 2013 20th IEEE International Conference on. IEEE, 2013:3137-3141.

[10]. Preetha R , Jinny S V . Early diagnose breast cancer with PCA-LDA based FER and neuro-fuzzy classification system[J]. Journal of Ambient Intelligence and Humanized Computing, 2021, 12(7):7195-7204.

[11]. Benouareth A . An efficient face recognition approach combining likelihood-based sufficient dimension reduction and LDA[J]. Multimedia Tools and Applications, 2021, 80(14):1-30.

[12]. Li J, Reddy Marpu P, Plaza A, et al. Generalized Composite Kernel Framework for Hyperspectral Image Classification[J]. IEEE Transactions on Geoscience & Remote Sensing, 2013, 51(9):4816-4829.

[13]. Li H , Jiang T , Zhang K . Efficient and robust feature extraction by maximum margin criterion[J]. IEEE Transactions on Neural Networks, 2006, 17(1):157-165.

[14]. Huang H, Yang M. Dimensionality Reduction of Hyperspectral Images With Sparse Discriminant Embedding[J]. IEEE Transactions on Geoscience & Remote Sensing, 2015, 53(9):5160-5169.

[15]. Luo H, Tang Y Y, Li C, et al. Local and Global Geometric Structure Preserving and Application to Hyperspectral Image Classification[J]. Mathematical Problems in Engineering, 2015, 2015:1-13.

[16]. Shi Q, Zhang L, Du B. Semisupervised Discriminative Locally Enhanced Alignment for Hyperspectral Image Classification[J]. IEEE Transactions on Geoscience & Remote Sensing, 2013, 51(9):4800-4815.