International Journal of Signal Processing, Image Processing and Pattern Recognition
Vol.9, No.12, (2016), pp.xx-xx
http://dx.doi.org/10.14257/ijsip.2016.9.12.27

Manifold Sparse Coding Based Hyperspectral Image Classification

Yanbin Peng[1*], Zhijun Zheng[1], Jiming Li[2], Zhigang Pan[1], Xiaoyong Li[1] and Zhinian Zhai[1]

[1]School of Information and Electronic Engineering, Zhejiang University of Science and Technology, Hangzhou 310023, China
[2]Department of Forensic Science, Zhejiang Police College, Hhangzhou 310053, China
pyb2010@126.com

Abstract

Hyperspectral image classification has received an increasing amount of interest in recent years. However, when representing pixels as vectors, the dimensionality of feature space is high, which causes "curse of dimensionality" problem. In this paper, in order to alleviate the impact of above problem, a manifold sparse coding method is proposed. Firstly, matrix decomposition technique is used to find a concept set and calculates relative data projection in the concept set. Secondly, manifold learning regularization is imported into objective function to capture the intrinsic geometric structure in the data. Finally, LASSO regularization is used to obtain sparse representation of data projection. Experimental results on real hyperspectral image show that the proposed method has better performance than the other state-of-the-art methods.

Keywords: *Manifold learning; sparse coding; hyperspectral image; pixel classification*

1. Introduction

In the research field of hyperspectral imaging, the remote sensors capture hyperspectral images in hundreds of narrow spectral bands. Pixels in hyperspectral image are represented as vectors. The reflection of spectral band corresponds to the entry in pixel vectors. The rich information contained in hyperspectral image data can provide accuracy and robust classification of the land-covers.

Recently, many machine learning methods have been developed to tackle the hyperspectral data classification problem. Camps-Valls [1] presented the kernel-based methods from a general viewpoint, and illustrates the main characteristics of different kernel approaches both theoretically and experimentally under the light of hyperspectral data classification. Chen [2] proposed a new algorithm for hyperspectral image classification based on sparse representation. A pixel is assumed to be sparsely represented by a few concepts in a given training dictionary. The sparse representation of a test spectral sample is recovered by solving a sparsity-constrained optimization problem via greedy pursuit algorithms. Kang [3] proposed a spectral-spatial classification framework based on edge-preserving filtering. Sun [4] proposed a novel task driven dictionary learning method with joint or Laplacian sparsity prior for hyperspectral image classification. The corresponding optimization algorithms are developed using fixed point differentiation, and are further simplified for ease of implementation. Li [5] considered a GMM classifier based on an LFDA- and LPNMF-induced feature subspace for hyperspectral image classification. The LFDA and LPNMF dimensionality-reduction

* Corresponding Author

ISSN: 2005-4254 IJSIP
Copyright ⓒ 2016 SERSC

techniques have superior locality-preserving properties and preserve the local manifold structure for hyperspectral data with complex distributions, from which the GMM classifier is able to accurately learn the class-conditional statistics. Di [6] presented a new sequential co-regularization active learning framework that utilizes multi-view consistency and the local proximity assumption for remote sensing image classification.

However, when representing pixels as vectors, the high number of spectral channels and low number of labeled training samples decrease the classification precision severely, which is called "curse of dimensionality" problem. To alleviate this problem, a dimensionality reduction step is usually adopted before classification. In this paper, in order to alleviate the impact of above problem, a manifold sparse coding method is proposed based on our former research works [7-10]. Firstly, matrix decomposition technique is used to find a concept set and calculates relative data projection in the concept set. Secondly, manifold learning regularization is imported into objective function to capture the intrinsic geometric structure in the data. Finally, LASSO regularization [11] is used to obtain sparse representation of data projection. Experimental results on real hyperspectral image show that the proposed method has better performance than the other state-of-the-art methods.

2. Hyperspectral Image Cube

Figure 1, shows the whole hyperspectral image cube H. I, J and K corresponds to three dimensions of the data cube. I and J stand for width and length dimension. K stands for the spectral dimension. One band image is represented as H_k, which is a data matrix with $I \times J$ dimensions. One pixel is represented as a vector x_i with K dimensions.

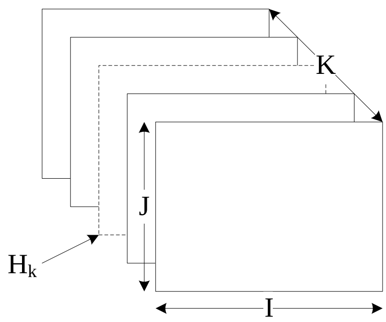

Figure 1. Hyperspectral Image Cube

3. Manifold Sparse Coding

In this section, we first state the general problem of Matrix decomposition technique, the objective function in the optimization process, and the traditional solution to the optimization problem. After this, the need of introducing manifold regularization in the objective function is presented. The formulation and characteristics of manifold regularization are briefly introduced. For a full theoretical description of manifold learning methods, the reader is referred to [12-14]. This section is closed with a LASSO problem to obtain sparse expression of hyperspectral image pixels.

3.1. Matrix Decomposition Technology

Pixel classification is the fundamental problem in hyperspectral image processing. Researchers have long sought efficient classification algorithm for pixels. For a given hyperspectral image, the pixel may has hundreds of distinct

 Copyright ⓒ 2016 SERSC

features. However, the freedom degree of each pixel could be far less. Instead of the original feature space, it is better to find a representative concept space to describe pixels. The dimensionality of data in representative concept space is much smaller than the original feature space. Matrix decomposition technique can be used to achieve this goal.

Given a data matrix $X=[x_1,x_2,...,x_n]\in R^{m\times n}$, each column of which corresponds to a pixel in hyperspectral image. Let $Q=[q_1,q_2,...,q_p]\in R^{m\times p}$ be the concept matrix, each column of which can be regarded as a basic concept. Let $R=[r_1,r_2,...,r_n]\in R^{p\times n}$ be the representation of original data in new concept space. Each column of R is the p-dimensional representation of the original pixels with respect to the concept set. Matrix decomposition technique is used to find these two matrixes Q and R so as to $X\approx QR$. Therefore, matrix decomposition technique can be regarded as a dimensionality reduction method since it reduces the dimension of pixels from m to p. The objective function of matrix decomposition can be formulated as follows:

$$\min_{Q,R}\|X-QR\|^2 \tag{1}$$

There are already lots of algorithms to solve matrix decomposition problem [15-16]. Different algorithms add different constraints on the above objective function to achieve different goals. LSA (Latent Semantic Analysis) [15] is a popular matrix decomposition algorithm. Based on SVD (Singular Value Decomposition), LSA requires the rank of matrix QR is less than k. NMF (Nonnegative Matrix Factorization) [16] is another popular matrix decomposition algorithm. Different form LSA, NMF requires that the entries in matrix Q and R are nonnegative.

3.2. Manifold Learning Regularization

Recently, researchers have considered that high-dimensional data, such as image, global climate patterns, or human gene expression, are sampled from a sub-manifold of the ambient Euclidean space. In fact, the pixel data in hyperspectral image can't fill up the high dimensional Euclidean space uniformly. Therefore, in the process of matrix decomposition, the intrinsic manifold structure should be considered to guide the dimensional reduction. In this paper, a manifold learning regularization is added into objective function of optimization process, with which we can calculate the concept set through an iterative computational method.

Given *n* data points $x_1,x_2,...,x_n$ where each data point corresponds to a pixel in hyperspectral image. We can construct a weighted graph with *n* nodes, one for each data point. The weight of edges connecting neighboring data points are defined as follows:

$$S_{ij}=\begin{cases} e^{-\frac{\|x_i-x_j\|^2}{t}} & if\ x_i\in KN(x_j)\ or\ x_j\in KN(x_i) \\ 0 & otherwise \end{cases} \tag{2}$$

Wherein $KN(x_i)$ denotes the set of *k* nearest neighbor data points of x_i. Parameter *t* is a real number. Define a diagonal matrix D, the entries of which are column sums of S. L=D-W is the Laplacian matrix. Manifold learning requires that the connected points in the weighted graph are as close as possible in the sub-manifold space. A reasonable method to achieve the requirement is to minimize the following function:

$$\frac{1}{2}\sum_{i,j}\| r_i - r_j \|^2 S_{ij} = trace(RLR^T) \tag{3}$$

Therefore, after adding the above manifold regularization, we can get the following objective function:

$$\min_{Q,R} \| X - QR \|^2 + \alpha \| Q \|^2 + \beta Trace(RLR^T) \tag{4}$$

Wherein, regularization term $\alpha \| Q \|^2$ is used to avoid over-fitting. $\beta \geq 0$ is the parameter of manifold regularization. To solve this optimization problem, we divide the algorithm into two steps: Learning projection matrix R while fixing the concept set Q, and learning concept set Q while fixing the projection matrix R.

3.2.1. Learning Projection Matrix R: This section discusses how to solve optimization problem (4) by fixing the concept matrix Q. the problem (4) becomes:

$$\min_{R} \| X - QR \|^2 + \beta Trace(RLR^T) \tag{5}$$

To solve problem (5), we update each column vector r_i in R individually, while fixing all the other column vectors $r_j (j \neq i)$ in R. The matrix decomposition optimization function $\| X - QR \|^2$ can be rewritten as follows:

$$\sum_{i=1}^{n} \| x_i - Qr_i \|^2 \tag{6}$$

The manifold learning regularization $Trace(RLR^T)$ can be rewritten as follows:

$$Trace(RLR^T) = Trace(\sum_{i,j=1}^{n} L_{ij} r_i r_j^T) = \sum_{i,j=1}^{n} L_{ij} r_j^T r_i = \sum_{i,j=1}^{n} L_{ij} r_i^T r_j \tag{7}$$

Therefore, the problem (5) can be rewritten as follows:

$$\min \sum_{i=1}^{n} \| x_i - Qr_i \|^2 + \beta \sum_{i,j=1}^{n} L_{ij} r_i^T r_j \tag{8}$$

We update r_i while letting the other vectors fixed. Thus, we get the optimization function:

$$\min f(r_i) = \| x_i - Qr_i \|^2 + \beta L_{ii} r_i^T r_i + 2\beta \sum_{j \neq i} L_{ij} r_i^T r_j \tag{9}$$

Let $\partial f(r_i) / \partial r_i = 0$, we can get the solution of value of r_i:

$$r_i = (Q^T Q + \beta L_{ii} I)^{-1} (Q^T x_i - \beta \sum_{j \neq i} L_{ij} r_j) \tag{10}$$

3.2.2. Learning Concept Matrix Q: This section discusses the method of learning concept matrix Q, while fixing the projection matrix R. The optimization problem (4) becomes:

$$\min_{Q} \| X - QR \|^2 + \alpha \| Q \|^2 \tag{11}$$

Copyright ⓒ 2016 SERSC

By making the derivative of Formula (11) with respect to Q and setting it to 0, we can get the optimal solution of matrix Q as follows:

$$Q = XR^T(\alpha I + RR^T)^{-1} \tag{12}$$

3.3. LASSO Regularization

The above section calculates the concept set matrix Q through importing manifold learning regularization term into matrix decomposition optimization function. To make the projection matrix R sparse, we import LASSO regularization to constrain the column of matrix R, letting most of the entries in column vector r_i becomes zero. The projection matrix R can be computed column by column independently through solving the following optimization problem.

$$\min_{r_i} \| x_i - Qr_i \|^2 + \delta | r_i | \tag{13}$$

Wherein x_i and r_i is the i-th column of matrix X and R respectively. LASSO adds the constraint $|r_i|$ into objective function to ensure the sparseness of r_i. The above optimization problem has the following equivalent formulation:

$$\begin{aligned} &\min_{r_i} \sum_{j=1}^{m} (x_{ji} - \sum_{k=1}^{p} q_{jk} r_{ki})^2 \\ &s.t.\ \sum_{k=1}^{p} | r_{ki} | \le \eta \end{aligned} \tag{14}$$

The pathwise coordinate optimization algorithm [17] can be used to solve the optimization problem in Equation (14). The bound η is a parameter, which is often chosen through a model selection procedure such as cross-validation. Problem (14) can also be transformed as a lagrange problem:

$$f(r_i) = \sum_{j=1}^{m} (x_{ij} - \sum_{k=1}^{p} q_{jk} r_{ki})^2 + \lambda \sum_{k=1}^{p} | r_{ki} | \tag{15}$$

Wherein $\lambda \ge 0$, λ in Formula (15) is corresponds to η in Formula (14). If $r_i(\lambda)$ is the optimization solution of Formula (15), then $\eta = \sum_{k=1}^{p} | r_{ki}(\lambda) |$.

When facing one dimensional projection problem, the lasso solution is simple. The sparse estimate of r_i is as follows [18]:

$$r_i^{lasso} = S(r_i^{'}, \lambda) = \begin{cases} r_i^{'} - \lambda & if\ r_i^{'} > 0\ and\ \lambda < | r_i^{'} | \\ r_i^{'} + \lambda & if\ r_i^{'} < 0\ and\ \lambda < | r_i^{'} | \\ 0 & if\ \lambda \ge | r_i^{'} | \end{cases} \tag{16}$$

Wherein $r_i^{'} = \sum_{j=1}^{m} x_{ji} q_j$ is the simple least-squares coefficient. Now we return to multi-dimensional problem. Assume the entries in r_i are uncorrelated, we can rewrite Formula (15) as follows:

$$f(r_i)=\sum_{j=1}^{m}(x_{ji}-\sum_{k\neq l}q_{jk}r_{ki}-q_{jl}r_{li})^2+\lambda\sum_{k\neq l}|r_{ki}|+\lambda|r_{li}| \tag{17}$$

Let all values of r_{ki} for $k\neq l$ fixed. Through Minimizing formula (17), we can get the solution of r_{li} as follows:

$$r_{li}=S(\sum_{j=1}^{m}q_{jl}(x_{ji}-\sum_{k\neq l}q_{jk}r_{ki}),\gamma) \tag{18}$$

4. Experimental Results

Having presented manifold sparse coding based method (MSC for short) in the previous sections, we now demonstrate the effect of our new method through a comparative experiment. The experiment is done in a real-world hyperspectral image data acquired by AVIRIS. The used hyperspectral image in our experiments was collected by the AVIRIS sensor over the Indian Pines region in 1992. Indian Pines data has 145 lines and 145 columns, which was acquired over a mixed agricultural and forest area. This data contains 220 spectral channels in the wavelength range from 0.4 to 2.5 micrometer. Spectral resolution is 10 nanometer. Spatial resolution is 20 meter. After pre-processing, several noise spectral bands were removed from the original data set, leaving a total of 200 channels to be used in our experiments. For illustrative purposes, Figure 2, shows the 30th band of the AVIRIS Indian Pines data. This data has 16 ground-truth classes.

Figure 2. 30th Band in India Pines Data Set

In order to assess MSC algorithm proposed in this paper, we choose two algorithms for comparison: 1) ID algorithm proposed in literature [19] with SVM classifier; 2) MVPCA algorithm proposed in literature [20] with SVM classifier.

Figure 3, displays the classification precision of algorithm ID, MVPCA and MSC. As is shown in Figure 3, the classification precision improves along with the increasing number of features. Among them, MSC algorithm achieves the best performance.

 Copyright ⓒ 2016 SERSC

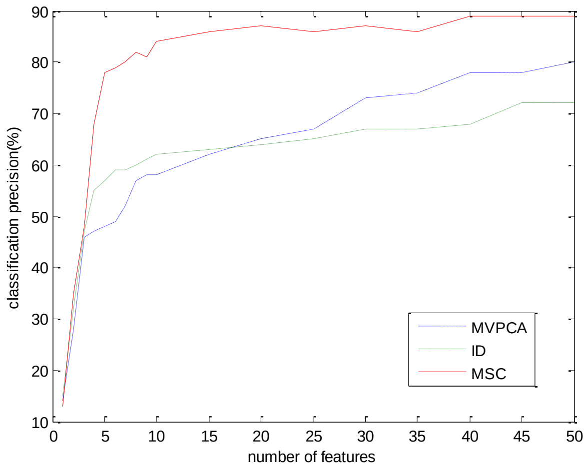

Figure 3. Comparison of Classification Precision

5. Conclusions

In this paper, we have presented a novel matrix decomposition method for hyperspectral image classification called manifold sparse coding (MSC for short). MSC have two steps to decrease the dimensional of data. In the first stage, through importing manifold learning regularization into objective function of matrix decomposition, MSC learns the concept set by exploiting the intrinsic geometric structure of the original data. In the second stage, MSC import LASSO method to learn a sparse representation with respect to the learned concept set for each pixel. Experimental results on real hyperspectral image show that the proposed method has better performance than the other state-of-the-art methods. In the future, on the one hand, manifold learning can be further exploited in the process of feature selection. On the other hand, we will apply transductive-SVM for the classification of hyperspectral image.

Acknowledgments

Project supported by the Zhejiang Provincial Natural Science Foundation of China (No. LY12F02018 & LQ13F020015).

References

[1] G Camps-Valls and L. Bruzzone, “Kernel-based methods for hyperspectral image classification”, IEEE Transactions on Geoscience and Remote Sensing, vol. 43, no. 6, **(2005)**, pp. 1351-1362.
[2] Y. Chen, N. M. Nasrabadi and T. D. Tran, “Hyperspectral image classification using dictionary-based sparse representation”, IEEE Transactions on Geoscience and Remote Sensing, vol. 49, no. 10, **(2011)**, pp. 3973-3985.
[3] X. Kang, S. Li and J. A. Benediktsson, “Spectral–spatial hyperspectral image classification with edge-preserving filtering”, IEEE Transactions on Geoscience and Remote Sensing, vol. 52, no. 5, **(2014)**, pp. 2666-2677.
[4] X. Sun, N. M. Nasrabadi and T. D. Tran, “Task-Driven Dictionary Learning for Hyperspectral Image Classification With Structured Sparsity Constraints”, IEEE Transactions on Geoscience and Remote Sensing, vol. 53, no. 8, **(2015)**, pp. 4457-4471.
[5] W. Li, S. Prasad and J. E. Fowler, “Hyperspectral image classification using Gaussian mixture models and Markov random fields”, IEEE Geoscience and Remote Sensing Letters, vol. 11, no. 1, **(2014)**, pp. 153-157.
[6] W. Di and M. M. Crawford, “Active learning via multi-view and local proximity co-regularization for hyperspectral image classification”, IEEE Journal of Selected Topics in Signal Processing, vol. 5, no. 3, **(2011)**, pp. 618-628.

Copyright ⓒ 2016 SERSC

[7] Z. ZHENG and Y. PENG, "Semi-Supervised Based Hyperspectral Imagery Classication", Sensors and Transducers, vol. 156, no. 9, **(2013)**, pp. 298-303.
[8] Y. Peng and J. Ai, "hyperspectral imagery classication based on spectral clustering band selection", Opto-Electronic Engineering, vol. 39, no. 2, **(2012)**, pp. 63-67.
[9] Z. Zheng and Y. Peng, "Spectral clustering and active learning based hyperspectral image classification", Journal of Computational Information Systems, vol. 11, no. 8, **(2015)**, pp. 2921-2928.
[10] Y. PENG and Z. ZHENG, "A Two-stage Sampling Based Ensemble Learning Method for Hyperspectral Image Classification", Journal of Computational Information Systems, vol. 11, no. 14, **(2015)**, pp. 5135-5142
[11] N. Simon, J. Friedman and T. Hastie, "A sparse-group lasso", Journal of Computational and Graphical Statistics, vol. 22, no. 2, **(2013)**, pp. 231-245.
[12] J. B. Tenenbaum, V. De Silva and J. C. Langford, "A global geometric framework for nonlinear dimensionality reduction", Science, vol. 290, no. 55, **(2000)**, pp. 2319-2323.
[13] M. Belkin and P. Niyogi, "Laplacian eigenmaps and spectral techniques for embedding and clustering", NIPS, **(2001)**, pp. 585-591.
[14] S. T. Roweis and L. K. Saul, "Nonlinear dimensionality reduction by locally linear embedding", Science, vol. 290, no. 55, **(2000)**, pp. 2323-2326.
[15] S. T. Dumais, "Latent semantic analysis", Annual review of information science and technology, vol. 38, no. 1, **(2004)**, pp. 188-230.
[16] C. J. Lin, "Projected gradient methods for nonnegative matrix factorization", Neural computation, vol. 19, no. 10, **(2007)**, pp. 2756-2779.
[17] J. Friedman, T. Hastie and H. Höfling, "Pathwise coordinate optimization", The Annals of Applied Statistics, vol. 1, no. 2, **(2007)**, pp. 302-332.
[18] D. L. Donoho and I. M. Johnstone, "Adapting to unknown smoothness via wavelet shrinkage", Journal of the american statistical association, vol. 90, no. 4, **(1995)**, pp. 1200-1224.
[19] C. I. Chang and S. Wang, "Constrained band selection for hyperspectral imagery", IEEE Transactions on Geoscience and Remote Sensing, vol. 44, no. 6, **(2006)**, pp. 1575-1585.
[20] C. I. Chang, Q. Du and T. L. Sun, "A joint band prioritization and band-decorrelation approach to band selection for hyperspectral image classification", IEEE Transactions on Geoscience and Remote Sensing, vol. 37, no. 6, **(1999)**, pp. 2631-2641.

Authors

Yanbin Peng, received his Ph.D. degree from the College of Computer Science and Technology, Zhejiang University. He has been associate professor at Zhejiang University of Science and Technology. His research interests include machine learning, image processing, hyperspectral image classification and unmixing.

 Copyright ⓒ 2016 SERSC