SVM classification of hyperspectral images based on wavelet kernel non-negative matrix factorization

Lin Bai^{*}, Meng Hui

School of Electronics and Control Engineering, Chang'An University, 710064, Xi'An, P.R.China

Received 1 March 2014, www.tsi.lv

Abstract

This paper presents a new kernel framework for hyperspectral images classification. In this paper, a new feature extraction algorithm based on wavelet kernel non-negative matrix factorization (WKNMF) for hyperspectral remote sensing images is proposed. By using the feature of multi-resolution analysis, the new method can improve the nonlinear mapping capability of kernel non-negative matrix factorization. The new classification method of hyperspectral image data combined with the novel kernel non-negative matrix factorization and support vector machine (SVM). The simulations results show that, the method of WKNMF reflect the nonlinear characteristics of the hyperspectral image. Experimental results on Airborne Visible Infrared Imaging Spectrometer 220 bands data in Indian pine test site and HYDICE 210 bands hyperspectral imaging in Washington DC Mall are both show that the proposed method achieved more strong analysis capability than comparative algorithms. Compared with the PCA, non-negative matrix factorization and kernel PCA method, classification accuracy of WKNMF with SVM can be improved over 5%-10%.

Keywords: hyperspectral, non-negative matrix factorization, classification, support vector machine, kernel method

1 Introduction

It is well know that each material has its own specific electromagnetic radiation spectrum characteristic. Using hyperspectral sensors, it is possible to recognize materials and their physical states by measuring the spectrum of the electromagnetic energy they reflect or emit. The spectral data which consist of hundreds of bands are usually acquired by a remote platform, such as a satellite or an aircraft, and all bands are available at increasing spatial and spectral resolutions. After 20 years of development, hyperspectral technology has not only been widely used in military, but also has been successfully applied in ocean remote sensing, vegetation surveys, geological mapping, environmental monitoring and other civilian areas [1, 2].

Due to the state of art of sensor technology developed recently, an increasing number of spectral bands have become available. Huge volumes of remote sensing images are continuously being acquired and archived. This tremendous amount of high spectral resolution imagery has dramatically increased the information source and increased the volume of imagery stored. For example, hyperspectral imagery captured by Airborne Visible Infrared Imaging Spectrometer (AVIRIS, operated by NASA) includes 224 bands, which contains up to 140Mbytes [2, 3].

However, the excessive hyperspectral data increase the difficulty of image processing and analysis. Such as supervised classification of hyperspectral images is a very challenging task due to the generally unfavourable ratio between the large number of spectral bands and the limited number of training samples available a priori, which results in the 'Hughes phenomenon'. Without the supports of new scientific concepts and novel technological methods, the existing large volumes of data prohibit any systematic exploitation. This has led to great demands to develop new concepts and methods to deal with large data sets [2-4].

Hyperspectral image classification has been a very active area of research in recent years [5]. Given a set of observations, the goal of classification is to assign a unique label to each pixel vector so that it is well-defined by a given class.

There are several important challenges when performing hyperspectral image classification. Supervised classification faces challenges related with the unbalance between high dimensionality and limited availability of training samples, or the presence of mixed pixels in the data. Another relevant challenge is the need to integrate the spatial and spectral information to take advantage of the complementarities that both sources of information can provide [5].

Over the last years, many feature extraction techniques have been integrated in processing chains intended for reduce the dimensionality of the data, thus mitigating the Hughes phenomenon. These methods can be unsupervised or supervised. Classic unsupervised techniques include principal component analysis (PCA), or independent component analysis (ICA). Supervised approaches comprise discriminant analysis for feature extraction (DAFE), decision boundary feature extraction

^{*} Corresponding author e-mail: Bai1981@sina.com.cn

(DBFE), and non-parametric weighted feature extraction (NWFE), among many others [4-7].

Recently, it was shown by Lee and Seung that positivity or non-negativity of a linear expansion is a very powerful constraint that also seems to yield sparse representations [8, 9]. Their technique, called nonnegative matrix factorization (NMF), was shown to be a useful technique in approximating high dimensional data where the data are comprised of nonnegative components. However, NMF and many of its variants are essentially linear, and thus can't disclose nonlinear structures hidden in the hyperspectral data. Besides, they can only deal with data with attribute values, while in many applications we do not know the detailed attribute values and only relationships are available. The NMF cannot be directly applied to such relation data. Furthermore, one requirement of NMF is that the values of data should be non-negative, while in many real world problems the non-negative constraints cannot be satisfied.

Since the mid-1990s, nuclear method has been successfully applied in the future, there are many scholars have proposed Nonlinear feature extraction method based on kernel method [10-13].

In this paper, a novel study is proposed for the feature extraction of high volumes of remote sensing images by using wavelet kernel non-negative matrix factorization (WKNMF). We propose the WKNMF, which can overcome the above limitations of NMF. Classification experiments on AVIRIS and HYDICE data sets by combination of feature extraction method and support the vector machine (SVM). The proposed method is applied to experiment data sets, compared with the other algorithms the classification accuracy can be increased over 5%-10%. The outline of this paper is as follows. Section 2 presents the proposed feature extraction based on WKNMF. Experimental results are reported in section 3. Finally, conclusions are given in section 4.

2 Methodology

2.1 NON-NEGATIVE MATRIX FACTORIZATION

NMF imposes the non-negativity constraints in learning the basis images. Both the values of the basis images and the coefficients for reconstruction are all non-negative. The additive property ensures that the components are combined to form a whole in the non-negative way, which has been shown to be the part based representation of the original data. However, the additive parts learned by NMF are not necessarily localized [8, 9].

Given the non-negative $n \times m$ matrix V and the constant r, the non-negative matrix factorization algorithm finds a non-negative $n \times r$ matrix W and another non-negative $r \times m$ matrix H such that they minimize the following optimality problem: min f(W, H).

Subject to
$$W \ge 0, H \ge 0$$
, (1)

Bai Lin, Hui Meng

This can be interpreted as follows: each column of matrix W contains a basis vector while each column of H contains the weights needed to approximate the corresponding column in V using the basis from W. So the product WH can be regarded as a compressed form of the data in V. The rank r is usually chosen $r \ll \min(n,m)$. f(W,H) is a loss function. In this paper, we set loss function as follow:

$$f(W,H) = \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{m} (V_{ij} - (WH)_{ij})^{2}.$$
 (2)

Solving the multiplicative iteration rule function as follows:

$$H_{bj} \leftarrow \frac{(W^T V)_{bj}}{(W^T W H)_{bj}}, W_{ib} \leftarrow W_{ib} \frac{(V H^T)_{ib}}{(W H H^T)_{ib}}.$$
(3)

The convergence of the process is ensured. The initialization is performed using positive random initial conditions for matrices W and H.

2.2 KERNEL NON-NEGATIVE MATRIX FACTORIZATION

Given *m* objects $\Theta_1, \Theta_2, \Theta_3, ..., \Theta_m$, with attribute values represented as an *n* by *m* matrix $\Omega = [\omega_1, \omega_2, ..., \omega_m]$, each column of which represent one of the m objects. Define the nonlinear map from original input space Ω to a higher or infinite dimensional feature space Φ as follows

$$\phi: x \in \Omega \to \phi(x) \in \Phi . \tag{4}$$

From the m objects, denote

$$\phi(\Omega) = [\phi(\omega_1), \phi(\omega_2), \dots, \phi(\omega_m)].$$
(5)

Similar as NMF, KNMF finds two non-negative matrix factors W_{ϕ} and H such that

$$\phi(\Omega) = W_{\phi}H \ . \tag{6}$$

 W_{ϕ} is the bases in feature space Φ and H is its combining coefficients, each column of which denotes now the dimension-reduced representation for the corresponding object. It is worth noting that since $\phi(\Phi)$ is unknown. It is impractical to directly factorize $\phi(\Omega)$. From Equation (6), we obtain

$$\left(\phi(\Omega)\right)^{T}\phi(\Omega) = \left(\phi(\Omega)\right)^{T} W_{\phi}H.$$
(7)

A kernel is a function in the input space and at the same time the inner product in the feature space through the kernel-induced nonlinear mapping. More specifically, a kernel is defined as

$$k(x, y) = \langle \phi(x), \phi(y) \rangle = (\phi(x))^T \phi(y) .$$
(8)

From Equation (8), the left side of Equation (7) can be rewritten as:

$$\begin{pmatrix} \phi(\Omega) \end{pmatrix}^T \phi(\Omega) = \left\{ (\phi(\omega_i))^T \phi(\omega_j) \right\}_{i,j=1}^m$$

$$= \left\{ k(\omega_i, \omega_j) \right\}_{i,j=1}^m = K,$$
(9)

Denote

$$Y = \left(\phi(\Omega)\right)^T W_{\phi} \,. \tag{10}$$

From Equation (9) and (10), Equation (7) can be changed as:

$$K = YH . (11)$$

Comparing Equation (11) with Equation (6), it can be found that the combining coefficient H is the same. Since W_{ϕ} is a learned base of $\phi(\Omega)$, similarly we call Yin Equation (11) as the bases of the kernel matrix K. Equation (11) provides a practical way for obtaining the dimension-reduced representation H by performing NMF on kernels.

For a new data point, the dimension-reduced representation is computed as follows

$$H_{new} = \left(W_{\phi}\right)^{+} \phi(\omega_{new})$$

= $\left(W_{\phi}\right)^{+} \left(\left(\phi(\Omega)\right)^{T}\right)^{+} \left(\phi(\Omega)\right)^{T} \phi(\omega_{new}).$ (12)
= $Y^{+}K_{new}$

Here A^+ donates the generalized (Moore-Penrose) inverse of matrix A, and $K_{new} = (\phi(\Omega))^T \phi(\omega_{new})$ is the kernel matrix between the m training instance and the new instance. Equation (11) and (12) construct the key components of KNMF when used for classification, it is easy to see that, the computing of KNMF need not to know the attribute values of objects, and only the kernel matrix K and K_{new} are required.

Obviously, KNMF is more general than NMF because the former can deal with not only attribute value data but also relational data. Another advantage of KNMF is that it is applicable to data with negative values since the kernel matrix in KNMF is always non-negative for some specific kernels.

2.3 WAVELET KERNEL NON-NEGATIVE MATRIX FACTORIZATION

The purpose of building kernel function is project hyperspectral observed data from low dimensional space to another high dimensional space. This WKNMF method uses the kernel function into the non-negative matrix factorization and improved it by replaced the traditional kernel function with the wavelet kernel function. By the feature of multi-resolution analysis, the nonlinear mapping capability of kernel non-negative matrix factorization method can be improved.

Assuming h(x) is a wavelet function, parameter α represent stretch and β represent pan. If there $x, x' \in \mathbb{R}^N$, then we get dot product form of wavelet kernel function:

$$K(x,x') = \prod_{i=1}^{N} h(\frac{x_i - \beta_i}{\alpha}) h(\frac{x'_i - \beta'_i}{\alpha}) .$$
(13)

Meet the reasonable expression product approved under the condition of translation invariance, the Equation (13) can be rewritten as:

$$K(x, x') = \prod_{i=1}^{N} h(\frac{x_i - x'_i}{\alpha}) .$$
 (14)

In this paper Morlet wavelet function was selected as generating function, according to the theory of translation invariance wavelet function, kernel function constructed as:

$$h(x) = \cos(1.75x)e^{(-x^2/2)}.$$
(15)

From Equation (13), (14) and (15) a wavelet kernel function meets the requirements of Mercer kernel function build as:

$$K(x,x') = \prod_{i=1}^{N} \left(\cos(1.75 \frac{(x_i - x_i')}{\alpha}) e^{\left(-\frac{\|x_i - x_i'\|^2}{2\alpha^2}\right)} \right).$$
(16)

Use Equation (16) in kernel non-negative matrix factorization, we can get Wavelet kernel non-negative matrix factorization.

2.4 SUPPORT VECTOR MACHINE CLASSIER INTRODUCTION

In machine learning, SVM are supervised learning models with associated learning algorithms that analyse data and recognize patterns, used for classification and regression analysis. The basic SVM takes a set of input data and predicts, for each given input, which of two

Bai Lin, Hui Meng

Bai Lin, Hui Meng

possible classes forms the output, making it a non-probabilistic binary linear classier.

Given a set of training examples, each marked as belonging to one of two categories, an SVM training algorithm builds a model that assigns new examples into one category or the other. The basic mathematical formula of SVM is:

$$\min_{w,b} \Phi(w,b) = \frac{1}{2} \|\omega\|^2$$

s.t.y_i(w·x_i-b) ≥ 1(i = 1,...,n). (17)

For more information about SVM see reference [14, 15].

3 Experimental results

3.1 EXPERIMENTAL ON AVIRIS DATA SET

The experiments were carried out on hyperspectral images produced by the AVIRIS. In order to simplify the logistics of marking this example analysis available to others, only a small portion of data set was chosen for this experiment. It contains 145 lines by 145 pixels (21025 pixels) and 190 spectral bands selected from a June 1992 AVIRIS data set of a mixed agriculture/forestry landscape in the Indian Pine Test Site in Northwestern Indiana.

For verification the feature extraction algorithm effect to hyperspectral data classification application, SVM classifier used in this paper. Given a set of training examples, each marked as belonging to one of two categories, an SVM training algorithm builds a model that assigns new examples into one category or the other. An SVM model is a representation of the examples as points in space, mapped so that the examples of the separate categories are divided by a clear gap that is as wide as possible. New examples are then mapped into that same space and predicted to belong to a category based on which side of the gap they fall on.

We select corn-min, corn-notil, soybean-min, soybean-notil and woods from AVIRIS images for classification experiment. Each object classes include 1434, 834, 968, 2468 and 1294 sample point respectively. The 3-bands (20, 80, 140 band) false colour synthesis image used in experiment and the ground truth are shown in Figure 1.



FIGURE 1 False colour images and ground truth of AVIRIS

Experiments using PCA, NMF, polynomial kernel KPCA (Poly-KPCK) comparison with WKNMF respectively, which Poly-KPCK coefficient kernel function is 5. To verify the classification capabilities of different feature extraction algorithm, we use Euclidean distance as the sum of the difference between the experimental data points in each band to take images of the same type of experimental data.

Take the Euclidean distance difference of surface features points between the different categories as a distance between the classes. The ratio of distance between the classes and distance within classes' values can reflect the degree to distinguish between different data. The experimental result was shown in Figure 2. From the experimental results, we can see WKNMF can get lowest ratio value than other algorithms. The result proves the new feature extraction method in this paper can effectively improve the discrimination between hyperspectral images category.



FIGURE 2 Ratio of distance between the classes and distance within classes

In order to verify the classification performance of feature extraction algorithm, experiments using the SVM method as a classifier, respectively PCA, NMF, polynomial KPCK (Poly-KPCA coefficient kernel function is 5) as feature extraction was compared with WKNMF. We use the overall accuracy (OA), as the evaluation index in experiment results.

Experiment randomly select 10% samples as training data on original hyperspectral data and the remaining 90% of sample as test data. The classification experiment

Bai Lin, Hui Meng

was repeated 10 times, taking the statistical average for final results.

Experiment with feature extraction algorithm, feature dimensions taken before 15 feature components as input, the energy of the total energy accounted for more than

TABLE 1 Classification results use 10% training sample data

96%. The classification result was shown as Table 1. An impact of feature dimensionality to the SVM classier for hyperspectral remote sensing images was shown as Figure 3.

Methods	corn-min	corn-notil	soybean-min	soybean-notil	woods	Total (OA)	Kappa
PCA	85.22	46.21	50.14	96.05	98.81	78.96	0.763
NMF	88.69	69.05	67.32	96.17	99.92	85.03	0.837
Poly-KPCA	88.13	73.54	65.89	97.11	99.92	86.81	0.849
WKNMF	92.93	82.12	76.88	96.61	99.81	91.91	0.893



FIGURE 3 Classification OAs with respect to reduced dimensionality in AVIRIS

The overall classification accuracy of test samples show that the use of a few training samples, WKNMF method can achieve higher classification accuracy. The classification accuracy enhance effect is very obvious, especially in corn-notil and soybean-min. Compared with the other algorithms WKNMF can improve the overall classification accuracy over 10%. We can see form table 1, variety of feature extraction algorithms not work very well in corn-notil and soybean-min, because of their spectral are similar and easily misclassification. Even under such adverse circumstances that the proposed method can still get higher classification accuracy than others.

3.2 EXPERIMENTAL ON HYDICE DATA SET

The Figure 4 shows a simulated colour IR view of an airborne hyperspectral data flightline over the Washington DC Mall provided with the permission of Spectral Information Technology Application Centre of Virginia who was responsible for its collection. The sensor system used in this case measured pixel response in 210 bands in the 0.4 to 2.4 μ m region of the visible and infrared spectrum. Bands in the 0.9 and 1.4 μ m region where the atmosphere is opaque have been omitted from the data set, leaving 191 bands. The data set contains 1208 scan lines with 307 pixels in each scan line. It totals

approximately 150 Megabytes. The image at left was made using bands 60, 27, and 17 for the red, green, and blue colours respectively. The HYDICE data set include Roofs, Street, Path (gravelled paths down the mall centre), Grass, Trees, Water, and Shadow.



FIGURE 4 False colour images of HYDICE

Experimental test data and training data are selected as shown in Table 2.

HYDICE data set(Washington DC Mall)							
classi	fication	samples					
Class No.	Class name	Train	Test				
1	Roofs	400	3434				
2	Street	168	248				
3	Path	36	139				
4	Grass	814	1114				
5	Trees	80	325				
6	Water	224	1000				
7	Shadow	11	86				

Bai Lin, Hui Meng

In order to verify the classification performance of feature extraction algorithm, experiments using the SVM method as a classifier, respectively PCA, NMF, polynomial KPCK (Poly-KPCA coefficient kernel function is 5) as feature extraction was compared with WKNMF. We use the overall accuracy (OA), as the evaluation index in experiment results. The classification experiment was repeated 10 times, taking the statistical average for final results. TABLE 3 Classification results on HYDICE data set

Experiment with feature extraction algorithm, feature dimensions taken before 20 feature components as input, the energy of the total energy accounted for more than 97%. The classification result was shown as Table 3. An impact of feature dimensionality to the SVM classier for hyperspectral remote sensing images was shown as Figure 5.

Class No.	Class name	Classification Algorithms				
		SVM	PCA+SVM	MNF+SVM	KPCA+SVM	WKNMF +SVM
1	Roofs	62.1%	64.8%	66.4%	70.7%	78.4%
2	Street	98%	100%	94.8%	98.4%	98.6%
3	Path	100%	100%	100%	100%	100%
4	Grass	97.2%	98.1%	97.7%	100%	99.8%
5	Trees	98.8%	98.8%	98.8%	95.4%	97.8%
6	Water	99.9%	99.9%	99.9%	99.8%	99.8%
7	Shadow	82.6%	79.1%	84.9%	89.5%	89.8%
overall accuracy		78.6%	80.7%	81%	84.1%	89.5%
Kappa		0.717	0.744	0.745	0.787	0.853



The overall classification accuracy of test samples show that the use of a few training samples, WKNMF method can achieve higher classification accuracy on HYDICE data set. Compared with the other algorithms WKNMF can improve the overall classification accuracy over 5%-10%.

4 Conclusions

In this paper, we propose a feature extraction of hyperspectral images by using WKNMF. The idea of using WKNMF techniques to find a set of basic functions to represent image data where the basic functions enable the identification and classification of intrinsic "parts" that make up the object being imaged by multiple observations. Experimental results on AVIRIS 220 bands data set in the Indian pine test site and HYDICE data sets in Washington DC Mall are both show that the proposed method achieved more strong analysis capability than comparative algorithms. Compared with the PCA, NMF and Poly-KPCA method, classification accuracy can be increased over 5%-10%. The WKNMF balance algorithm efficiency and performance very well.

Acknowledgments

The project was supported by national natural science foundation (No.41101357) and the special fund for basic scientific research of central colleges, Chang'an University, number CHD2011JC170 and CHD2011TD018.

References

- [1] Muñoz-Marí J, Tuia D, Camps-Valls G 2012 Semisupervised classification of remote sensing images with active queries IEEE Trans. Geosci. Remote Sensing 50(10) 3751-63
- [2] Bioucas-Dias J M, Plaza A, Scheunders C-V G, Nasrabadi P, Chanussot N M 2013 Hyperspectral Remote Sensing Data Analysis and Future Challenges Geoscience and Remote Sensing Magazine IEEE 1(2) 6-36
- [3] Tuia D, Volpi M, Copa L, Kanevski M, Muñoz-Marí J 2011 A survey of active learning algorithms for supervised remote sensing image classification IEEE J. Select. Topics Signal Processing 5(3) 606 - 17
- [4] Dopido, I.Villa, A., Plaza, A., Gamba, P. 2012 A Quantitative and Comparative Assessment of Unmixing-Based Feature Extraction Techniques for Hyperspectral Image Classification IEEE Journal of selected topics in applied earth observations and remote sensing **5**(2), pp 421-435
- [5] José M. Bioucas-Dias, Antonio Plaza, G. Camps-Valls, Paul Scheunders, Nasser Nasrabadi, and Jocelyn Chanussot 2013 Hyperspectral Remote Sensing Data Analysis and Future Challenges IEEE Geoscience and Remote Sensing Magazine 1(2) 6-36
- [6] Liu F, Gong J Y 2009 A classification method for high spatial resolution remotely sensed image based on multi-feature Geography and Geo-Information Science 25(3) 19-41
- [7] Su H, Yang H, Du Q, Sheng Y H 2011 Semisupervised band clustering for dimensionality reduction of hyperspectral imagery IEEE Geoscience and Remote Sensing Letters 8(6) 1135-9
- [8] Sen Jia, Yuntao Qian 2009 Constrained Nonnegative Matrix Factorization for Hyperspectral Unmixing IEEE Transactions on Geoscience and Remote Sensing 47(1) 161-73

- [9] Xuesong Liu, Wei Xia, Bin Wang, Liming Zhang 2011 An approach based on constrained nonnegative matrix factorization to unmix hyperspectral data IEEE Transactions on Geoscience and Remote Sensing 49(2) 757-72
- [10]Song Xiangfa, Jiao Licheng 2012 Classifiction of hyperspectral remote sensing image based on sparse representation and spectral information Journal of Electronics&Information Technology 34(2) 268-73
- [11] Fauvel M, Chanussot J, Benediktsson J A 2012 A Spatial spectral Kernel based Approach for the Classification of Remote sensing Images Pattern Recognition 45(1) 381-92
- [12] Chen Yi, Nasser M, Trac D T 2013 Hyperspectral Image Classification via Kernel Sparse Representation IEEE Transactions on Geoscience and Remote Sensing 51(1) 217-31
- [13] Wang J, Jabbar M A 2012 Multiple Kernel Learning for adaptive graph regularized nonnegative martix factorization national association of science and technology for development 41(3) 115-22
- [14] Yushi Chen, Xing Zhao, Zhouhan Lin 2014 Optimizing Subspace SVM Ensemble for Hyperspectral Imagery Classification IEEE Journal of Selected Topics in Applied Earth Observations and *Remote Sensing* **7**(4) 1295-305
- [15] Bor-Chen Kuo, Hsin-Hua Ho, Cheng-Hsuan Li, Chih-Cheng, Hung A 2014 Kernel-Based Feature Selection Method for SVM With RBF Kernel for Hyperspectral Image Classification IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 7(1) 317-26

Authors

Lin Bai, born on August 16, 1981, Shannxi, PR.China Current position, grades: Lecturer University studies: 1999/09-2003/06 Northwestern University gets BS. degree. Major: Electronic Information Science and Technology. 2006/09-2006/03 Xidian University gets MS. degree. Major: Circuits and Systems. 2006/03-2011/07 Northwestern Polytechnical University get Ph.D. degree. Major: Signal and information processing Scientific interest: Signal Processing, Image Processing, Remote sensing information processing. Publications: The author has published over 10 papers including 6 in El index and 1 in SCI index. Experience: 2011/07- Chang'an University, school of Electronic and Control Engineering. MengHui, born on October 19, 1981, Shannxi, PR. China Current position: Lecturer University studies: 2000/09-2004/07Xi'an JiaoTong University gets BS. degree. Major: Mechanical Engineering. 2004/09-2007/01Xi'an JiaoTong University gets MS. degree. Major: Mechanical Engineering. 2006/03-2011/07 Xi'an JiaoTong University get Ph.D. degree. Major: Electrical Engineering. Scientific interest: Nonlinear system, Chaos, Electrical Engineering. Publications: The authors has published 14 papers including 6 papers in SCI index. Experience: 2012/04 - Chang'an University, school of Electronic and Control Engineering.