A Nonlinear Regression Classification Algorithm with Small Sample Set for Hyperspectral Image

Jiayi Li, Hongyan Zhang, and Liangpei Zhang*

The State Key Laboratory of Information Engineering in Surveying, Mapping, and Remote Sensing, Wuhan University, P.R. China

*Corresponding author: zhanghongyan@whu.edu.cn

ABSTRACT

A column generation kernel technology based nonlinear regression classification method for hyperspectral image is proposed in this paper. The nonlinear extension for the collaborative representation regression is utilized in the joint collaboration model framework. The proposed algorithm is tested on two hyperspectral images. Experimental results suggest that the proposed nonlinear algorithm shows superior performance over other linear regression-based algorithms and the classical hyperspectral classifier SVM.

Index Terms—column generation, collaborative representation, hyperspectral image classification, kernel

1. INTRODUCTION

Hyperspectral image (HSI), spanning the visible to infrared spectrum with hundreds of continuous narrow spectral bands, can facilitate discrimination of object types. Meanwhile, obstacles such as nonlinear correlations between the higher-order spectral inter-band and lack of available training samples etc, appear as spectral resolution and data dimensionality increase. In view of this, supervised classification of high dimensional data set with small sample set is still a difficult endeavor.

In recent years, a novel collaborative linear regression approach for recognition has been introduced into high-dimensional classification tasks [1], where the usage of collaborative representation (CR) as an effective mechanism leads to state-of-the-art performance. The CR technique has also been applied to HSI classification [2], relying on the observation that hyperspectral test pixel can be approximately represented by a given dictionary constructed from training samples. With aid from the rest training samples, the CR-based classifier can work efficiently in the lack of sample case [1]. In view of this approach, we further extend nonlocal joint collaborative representation classification (NJCRC) [2] into a nonlinear version with a column generation (CG) kernel technology [3]. This method firstly maps the origin spectral space to a higher kernel space by directly taking the similar measures between spectral pixels as new feature, and then utilizes nonlocal joint collaborative regression model for kernel signal reconstruction and sequential pixel classification. Unlike the kernel trick used in various approaches [4], the CG-strategy is easy to implement and do not require the explicit inner product structure in the regression analytical solution.

The remaining part of this paper is organized as follows. Details about the proposed KNJCRC algorithm are described in Section 2. Section 3 shows the experimental results conducted on several HSIs with the proposed algorithm and several state-of-the-art classification methods. Finally, Section 4 summarizes our work.

2. CLASSIFICATION OF HSI USING COLLABORATIVE REPRESENTION

In this section, we firstly briefly introduce the CR-based algorithms for HSI classification, and then extend the linear version algorithms into the column generation kernel space, in which these hyperspectral classes will be linearly separable.

2.1. Collaborative representation classification

For collaborative representation classification (CRC) [1], suppose we have M distinct classes and N_i $(i = 1, \ldots, M)$ training samples for each class. In the classical collaborative representation model, training samples from the ith class as columns of a sub-dictionary $\mathbf{A}_i = \left[\boldsymbol{a}_{i,1}, \boldsymbol{a}_{i,2}, \ldots \boldsymbol{a}_{i,N_i} \right] \in \mathbb{R}^{B \times N_i}$, then the

collaborative dictionary $\mathbf{A} \in \mathbb{R}^{B \times N}$ with $N = \sum_{i=1}^{M} N_i$ is constructed by combining all the sub-dictionaries $\{\mathbf{A}_m\}_{m=1,\ldots,M}$. Thus, an unknown test pixel $\boldsymbol{s} \in \mathbb{R}^B$ can be written as a collaborative linear combination of all of the training samples as

$$\boldsymbol{s} = \mathbf{A}\boldsymbol{\alpha} + \boldsymbol{\varepsilon} = \mathbf{A}_i\boldsymbol{\alpha}_i + \sum_{j=1, j\neq i}^{M} \mathbf{A}_j\boldsymbol{\alpha}_j + \boldsymbol{\varepsilon} \in \mathbb{R}^B \tag{1}$$

where $\boldsymbol{\varepsilon}$ is a small constant for noise of the signal. The collaborative coefficient vector $\boldsymbol{\alpha}$ can be obtained by solving the following optimization problem:

$$\hat{\boldsymbol{\alpha}} = \arg\min_{\boldsymbol{\alpha}} \left\{ \left\| \boldsymbol{s} - \mathbf{A}\boldsymbol{\alpha} \right\|_2 + \lambda \left\| \boldsymbol{\alpha} \right\|_2 \right\} \tag{2}$$

The classification rule for CRC via regularized least squares which referred to as CRC-RLS [1] is denoted :

$$class(\boldsymbol{s}) = \arg\min_{i=1,\ldots,M} \left\| \boldsymbol{s} - \mathbf{A}_i\hat{\boldsymbol{\alpha}}_i \right\|_2 / \left\| \hat{\boldsymbol{\alpha}}_i \right\|_2 \tag{3}$$

2.2. Nonlocal joint collaborative representation classification

For hyperspectral data, pixels in a small neighborhood with similar spectrum can be represented in a same low-dimensional feature subspace by different compact coefficients. In a spatial patch centered at spatial position c, we nonlocally select K most similar pixel with the central pixel $\boldsymbol{s}_c$ by KNN method [5], and these K pixel can be stacked as $\mathbf{S}_K = \left[\boldsymbol{s}_1, \ldots, \boldsymbol{s}_K\right] \in \mathbb{R}^{B \times K}$. It is believed that these K pixels share a "common collaboration pattern" as they are selected by the measure of the correlations between the central test pixel $\boldsymbol{s}_c$, not the spatial distance. In this context, the joint signal matrix can be represented as:

$$\begin{aligned} \mathbf{S}_K &= [\boldsymbol{s}_1\ \boldsymbol{s}_2\ \ldots \boldsymbol{s}_K] = \mathbf{A}\left[\boldsymbol{\alpha}_1\ \boldsymbol{\alpha}_2\ \ldots \boldsymbol{\alpha}_K\right] + \boldsymbol{\Sigma} \\ &= \mathbf{A}_i\boldsymbol{\Psi}_{K(i)} + \sum_{j=1, j\neq i}^{M} \mathbf{A}_j\boldsymbol{\Psi}_{K(j)} + \boldsymbol{\Sigma} = \mathbf{A}\boldsymbol{\Psi}_K + \boldsymbol{\Sigma} \end{aligned} \tag{4}$$

where $\boldsymbol{\Psi}_K$ is a set of all the coefficient vectors , $\boldsymbol{\Psi}_{K(i)}$ is the ith subset of $\boldsymbol{\Psi}_K$ and $\boldsymbol{\Sigma}$ is the random noise matrix corresponding to the joint signal matrix. By the CR-based approach, (4) can be calculated as

$$\hat{\boldsymbol{\Psi}}_K = \arg\min_{\boldsymbol{\Psi}_K} \left\{ \left\| \mathbf{S}_K - \mathbf{A}\boldsymbol{\Psi}_K \right\|_F^2 + \lambda \left\| \boldsymbol{\Psi}_K \right\|_F^2 \right\} \tag{5}$$

The label of the center pixel $\boldsymbol{s}_c$ is determined by the following classification rule:

$$class(\boldsymbol{s}_c) = \arg\min_{i=1,\ldots,M} \left\{ \left\| \mathbf{S}_K - \mathbf{A}_i\hat{\boldsymbol{\Psi}}_{K(i)} \right\|_F / \left\| \hat{\boldsymbol{\Psi}}_{K(i)} \right\|_F \right\} \tag{6}$$

where $\mathbf{A}_i$ is a sub-part of $\mathbf{A}$ in class i, and $\hat{\boldsymbol{\Psi}}_{K(i)}$ denotes the corresponding portion of the recovered collaborative coefficients in the i th class.

2.3. Column generation kernel technology

Column generation is widely used in linear programming since the 1950s. The kernel mapping used in this paper, which is similar to a simplified column generation strategy for CG-Boost [3] in multiple kernel leaning, directly take the signal in kernel space as feature [5] .

For the kernel function, the real-value function $\kappa : \mathbb{R}^B \times \mathbb{R}^B \mapsto \mathbb{R}$ is defined as the inner product: $\kappa\left(\boldsymbol{x}_i, \boldsymbol{x}_j\right) = \left\langle \phi(\boldsymbol{x}_i), \phi(\boldsymbol{x}_j) \right\rangle$, where $\phi(\boldsymbol{x})$ is a function of spectral vector $\boldsymbol{x}$. In this paper, we utilize the RBF kernel:

$$\kappa\left(\boldsymbol{x}_i, \boldsymbol{x}_j\right) = \exp\left(-\gamma\rho\left(\boldsymbol{x}_i, \boldsymbol{x}_j\right)\right) \tag{7}$$

where $\gamma > 0$ controlling the width of the RBF kernel, and ρ define a "distance" between $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$. The chi-square distance $\chi^2\left(\boldsymbol{x}_i, \boldsymbol{x}_j\right)$, which can reflect the relative difference between corresponding spectral sub-region are used in this paper.

$$\kappa\left(\boldsymbol{x}_i, \boldsymbol{x}_j\right) = \exp\left(-\chi^2\left(\boldsymbol{x}_i, \boldsymbol{x}_j\right) / \mu\right) \tag{8}$$

where μ is set to the mean value of pairwise chi-square distance and is adaptive to the training set .

In this paper, denote $\boldsymbol{s} \in \mathbb{R}^B$ as the data point of interest and $\boldsymbol{s}' \in \mathbb{R}^N$ as its representation in the feature space. The kernel collaborative representation of test pixel $\boldsymbol{s}$ in terms of all training pixels $\boldsymbol{a}_i$'s can be formulated as

$$\begin{aligned} \boldsymbol{s}' &= \left[\kappa\left(\boldsymbol{a}_1, \boldsymbol{s}\right) \cdots \kappa\left(\boldsymbol{a}_N, \boldsymbol{s}\right)\right]^T \\ &= \underbrace{\begin{pmatrix} \kappa\left(\boldsymbol{a}_1, \boldsymbol{a}_1\right) \cdots \kappa\left(\boldsymbol{a}_1, \boldsymbol{a}_N\right) \\ \vdots \\ \kappa\left(\boldsymbol{a}_N, \boldsymbol{a}_1\right) \cdots \kappa\left(\boldsymbol{a}_N, \boldsymbol{a}_N\right) \end{pmatrix}}_{\kappa(\mathbf{A})} \underbrace{\left[\alpha_1' \cdots \alpha_N'\right]^T}_{\boldsymbol{\alpha}'} \\ &= \kappa\left(\mathbf{A}\right)\boldsymbol{\alpha}' \end{aligned} \tag{9}$$

where the columns of $\kappa(\mathbf{A})$ are the representation of training samples in the feature space, and $\boldsymbol{\alpha}'$ is assumed to be a $N \times 1$ kernel representation vector.

2.4. Kernel nonlocal collaborative representation classification via column generation technology

For the nonlocal joint representation model, we can also extend the $\mathbf{S}_K \in \mathbb{Z}^{B \times K}$ into the feature space. We first map all the pixels in the spatial window sized $T(T \geq K)$ into the kernel feature space, then calculate the correlation between every kernel signal in the neighborhood window and the test kernel signal $\boldsymbol{s}_c'$, and finally sort all kernel signals in the order of descending correlation. We select the first K ones from all the T kernel signals and consider them also share a "common collaborative pattern" in the feature space. The nonlocal kernel joint signal matrix can be represented in the kernel feature space as:

$$\mathbf{S}_K' = \kappa(\mathbf{A}) \underbrace{\begin{pmatrix} \alpha_{1,(1)}' \cdots \alpha_{1,(K)}' \\ \vdots \\ \alpha_{N,(1)}' \cdots \alpha_{N,(K)}' \end{pmatrix}}_{\boldsymbol{\Psi}_K'} = \kappa(\mathbf{A}) \boldsymbol{\Psi}_K' \tag{10}$$

where $\boldsymbol{\Psi}_K'$ is the kernel collaborative coefficient matrix, and the function (6) can be extended as :

$$\hat{\boldsymbol{\Psi}}_K' = \arg\min_{\boldsymbol{\Psi}_K'} \left\{ \left\| \mathbf{S}_K' - \kappa(\mathbf{A}) \boldsymbol{\Psi}_K' \right\|_F^2 + \lambda \left\| \boldsymbol{\Psi}_K' \right\|_F^2 \right\} \tag{11}$$

Once the coefficient matrix $\hat{\boldsymbol{\Psi}}_K'$ is obtained, the classification rule is denoted as:

$$class(\boldsymbol{s}_c) = \arg\min_{i=1,\ldots,M} \left\| \mathbf{S}_K' - \kappa(\mathbf{A}_i) \hat{\boldsymbol{\Psi}}_{K,i}' \right\|_F^2 / \left\| \hat{\boldsymbol{\Psi}}_{K,i}' \right\|_F^2 \tag{12}$$

where $\kappa(\mathbf{A}_i)$ is a sub-part of $\kappa(\mathbf{A})$ in class i, and $\hat{\boldsymbol{\Psi}}_{K,i}'$ denotes the portion of the recovered kernel collaborative coefficients corresponding to the entire training samples in the i th class.

3. EXPERIMENTAL RESULTS AND ANALYSIS

In this section, we demonstrate the effectiveness of the proposed algorithm on two hyperspectral images. The classical classifier SVMs with RBF kernel [4], JSRC [7] and NJCRC [2] are used as benchmarks in this paper.

The first hyperspectral image in this paper was gathered by AVIRIS sensor over the Indian Pines test site in North-western Indiana and consists of pixels and 224 spectral reflectance bands in the wavelength range 0.4–2.5×10^(-6) meters. We have also reduced the number of bands to 200 with 24 water absorption bands being removed. The false color image is visually shown in Figure 1(a). This image contains 10 ground truth classes which can be visually shown in Figure 1(b). In this experiment, we randomly sample 60 pixels of the data in each class as the training samples and the remaining as the test samples, and the detailed information is shown in Table 1. The classification accuracy for each class using different classifiers is also shown in Table 1 and the classification maps are shown in Figure 1(c)-(f). The residual and optimal tolerance parameters for each greedy algorithm mentioned above are default. The optimal parameters for the KNJCRC are $\lambda = 1e-7$, $T = 81$ and $K = 50$, for the NJCRC are $\lambda = 1e-5$, $T = 81$ and $K = 55$, where the corresponding optimal neighboring size for JSRC is $T = 25$. Besides, the parameters for SVM are obtained by 10 fold cross-validation. It is shown from Table 1 that by mapping the original spectral feature into a higher kernel feature space, the proposed algorithm outperforms the other classification algorithms, and is superior to its linear version algorithm.

The second hyperspectral image used in this paper is the 115-band ROSIS image Centre of Pavia of size 776×485, for which we use only 102 bands with the 13 water absorption bands removed. The false color image can be visually shown in Figure 2(a). There are 9 classes of interests as shown by the ground truth map in Figure 2(b), in this experiment, we randomly sample 20 pixels per class of the data as the training samples and the remaining as the test samples, and the detailed is shown in Table 2 which contains the classification accuracy for each class using various and the classification maps are shown in Figure 2(c)-(f). It is observed that we can draw the same conclusion with the first experiment.

4. CONCLISIONS

In this paper, we propose a new HSI classification technique based on collaborative representation in a nonlinear feature space induced by column generation kernel method. The nonlocal contextual correlation is incorporated to constrain the dominated representation through the joint collaboration representation. The kernel technique in this paper is different from the conventional kernel mapping in RKHS feature space by kernel trick. The column generation directly treats the similar

measures between spectral pixels as feature, while the conventional kernel method often replaces the original feature vector as implicit kernel feature by the inner product operation. The extensive experimental results clearly suggest the proposed method can achieve competitive classification results. Our further work will focus on more brilliant contextual information extraction which automatic obtain the joint signal matrix to further improve the classification performance.

5. REFERENCES

[1] L. Zhang, M.Yang, X. Feng, Y. Ma, and D. Zhang,"Collaborative Representation based Classification for Face Recognition,"*Arxiv preprint arXiv:1204.2358,* 2012.

[2] J. Li, H. Zhang, L. Zhang and Y. Huang., "Hyperspectral Image Classification by Nonlocal Joint Collaborative Representation with Locality-adaptive Dictionary," *IEEE Trans. Geosci. Remote Sens.*, *submitted*

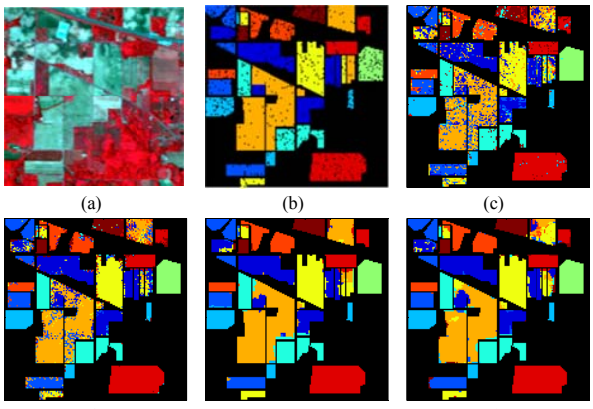
[3] J. Bi, T. Zhang and K.P. Bennett., "Column-generation boosting methods for mixture of kernels," *in Conference on Knowledge Discovery in Data: Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining,* 2004, pp. 521-526.

[4] C.I. Chang, "Hyperspectral imaging: Techniques for spectral detection and classification", *Springer Us*, vol. 1, 2003.

[5] X. Yuan, X. Liu, and S. Yan, "Visual Classification with Multi-task Joint Sparse Representation," *IEEE Trans. Image Process.*, *to appear*.

[6] H. Zhang, A.C. Berg, M. Maire, and J. Malik, "SVM-KNN: Discriminative Nearest Neighbor Classification for Visual Category Recognition,"in *Proc. IEEE Computer Society Conf.Computer Vision and Pattern Recognition*, 2006, pp. 2126–2136.

[7] Y. Chen, N. M. Nasrabadi, and T. D. Tran, "Hyperspectral Image Classification Using Dictionary-Based Sparse Representation," *IEEE Trans. Geosci. Remote Sens.,* vol. 49, no. 10, pp. 3973–3985, Oct. 2011.

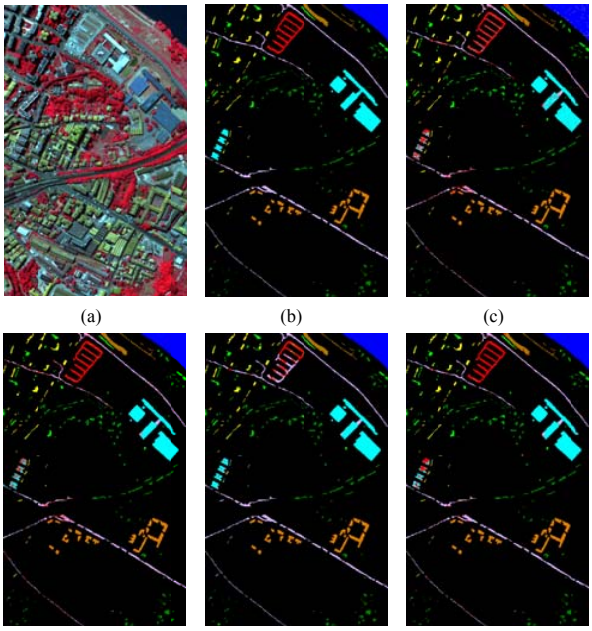


(a) (b) (c) (d) (e) (f)

Figure 1.Classification results of Indian Pines image: (a) false color image (R:57 G:27 B:17) (b) ground truth (c) SVM, (d) JSRC, (e) NJCRC, (f) KNJCRC

Table 1. Classification accuracy(%) for the Indiana Pine image on the test set using different classifiers

Class	Train	Test	SVM	JSRC	NJCRC	KNJCRC
1	60	1368	0.6923	0.8509	**0.9393**	0.8882
2	60	770	0.7792	0.9091	0.9416	**0.9948**
3	60	423	0.9125	0.9409	0.9173	**0.9456**
4	60	670	0.9418	**0.9970**	**0.9970**	0.9910
5	60	418	0.9952	**1**	**1**	**1**
6	60	912	0.7160	0.7971	**0.9748**	0.9276
7	60	2395	0.5908	0.6731	0.7779	**0.8342**
8	60	533	0.7992	0.8180	**0.9681**	0.9306
9	60	1205	0.9245	0.9593	0.9743	**0.9876**
10	60	326	0.7178	0.9479	**1**	0.9969
OA	600	9020	0.7563	0.8412	0.9149	**0.9222**
Kappa			0.7203	0.8177	0.9018	**0.9100**



(a) (b) (c) (d) (e) (f)

Figure 2.Classification results of Centre of Pavia image: (a) false color image, (R:102, G:56 ,B:31), (b) ground truth (c) SVM, (d) JSRC, (e)NJCRC, (f) KNJCRC

Table 2. Classification accuracy(%) for the Centre of Pavia image on the test set using different classifiers

Class	Train	Test	SVM	JSRC	NJCRC	KNJCRC
1	20	5290	0.9809	**1**	**1**	**1**
2	20	3486	0.8448	0.8890	**0.9317**	0.8964
3	20	958	0.9614	**0.9676**	0.8904	0.9259
4	20	2120	0.8542	0.9443	0.5170	**0.9920**
5	20	1069	0.8877	0.9476	**1**	0.9355
6	20	4869	0.8316	0.7989	**0.9881**	0.9499
7	20	7267	0.8563	0.9180	**0.9622**	0.9040
8	20	3102	0.9700	0.9923	**1**	0.9997
9	20	1599	**0.9950**	0.9287	0.7448	0.7967
OA	180	29940	0.8967	0.9225	0.9292	**0.9400**
Kappa			0.8793	0.9092	0.9162	**0.9295**

6. ACKNOWLEDGEMENTS

This work was supported in part by the National Basic Research Program of China (973 Program) under Grant 2011CB707105, by the National Natural Science Foundation of China under Grants 61201342 and 40930532 and by LIESMARS Special Research Funding.