Class-Probability Based Semi-supervised Dimensionality Reduction for Hyperspectral Images

Lu Liang
School ojElectrical Engineering and Automation
Anhui University
Hejei, Anhui Province, China
Email: lianglul99I@I26.com

Yi Xia*, Lina Xun
School ojElectrical Engineering and Automation
Anhui University
Hejei, Anhui Province, China
*Correspondence: xiayi@ahu.edu.cn

Qing Yan, Dexiang Zhang
School ojElectrical Engineering and Automation
Anhui University
Hejei, Anhui Province, China
Email: dqxyzdx@I26.com

Abstract—**Hyperspectral** images (HSI) are very useful due to the rich information they contained However, for the same reason, it is also inconvenient to be analyzed due to its high dimension and also because it contains a lot of redundant lnformation, Therefore, dimensionality reduction (DR) is often an lndispensable step for the analysis of HSI. Due to the expensiveness of labeling samples, semi-supervlsed learning technique that performs DR with only a small amount of labeled samples, has attract more and more attention during the past several years. In **this** paper, we propose a novel method called class prebablllty semi-supervised DR (CPSDR). Unlike previously semi-supervlsed DR methods, which only focus on a small number of labeled samples and depend on their local geometry information, our approach also pay much attention on unlabeled samples. Moreover, in our approach, not only local geometry information but also class structure information was exploited. We then combined these two information together to yield a more dlscrtmlnatlve scatter matrix. We formulate ow' problem as an optlmizatlon problem and solve it by eigenvalue decomposition. The experimental results on Salinas and PaviaU hyperspectral data suggested that our algorithm achieved state-of-the-art performance.

Keywords-dimensionality reduction; class probability; semi-supervised; hyperspectral image

I. INTRODUCTION

Hyperspectral image (HSI) [1,2] contains abundant information which sampled from a continuous narrow spectral band. With the rapid development of geostationary observation technology, the spectral resolution of the hyperspectral image has been improved substantially, leading to wide applications such as detection, identification, clustering, classification and so on. However, due to the high dimension of HSI data, which may cause the "curse of dimensionality" as well as computational complexity. So, dimensionality reduction (DR) is often used in HSI to alleviate the "information rich and knowledge poor" status and reduce the complexity of data [3]. Hence, DR becomes a crucial step in many practical machine learning tasks.

Recently, the graph based semi-supervised learning methods [4] has attracted more attention, because they can get better performance with less labeled information. Most of the supervised and semi-supervised algorithms that have been proposed can't guarantee their performance if the labeled data samples are insufficient. In addition, the large amount of information contained in unlabeled data has not been given enough attention, and most semi-supervised methods [5-7] simply utilized labeled samples as training data. In recent years, under the assumption that adjacent samples has much possibility to be with same label, some supervised DR techniques [4] also include the unlabeled samples during the optimization process of the objective function.

In this study, the labeled and unlabeled data are given equally important roles in the formulation of our DR problem. Our proposed method, called as class probability semi-supervised DR (CPSDR), explored not only the class structure information of unlabeled data but also the local information of the labeled data. Experiments indicates that our approach can form more discriminative intra-class and inter-class scatter matrix, resulting in state-of-the-art classification performance.

The remainder of the paper is organized as follows. Section II gives a brief review of probabilistic semi-supervised discriminant analysis (PSDA) DR methods. Our proposed algorithm is described carefully in Section III. Then, the experimental results of our proposed method together with several state-of-the-art DR algorithms on two real hyperspectral data sets are reported in Section IV Finally, Section V concludes this paper.

II. RELATED W ORK

Assuming that a partially labeled hyperspectral data set $X = [x_1, x_2, \cdots, x_N] \in R^{D \times N}$ with N samples, where B represents the number of bands or the dimension for hyperspectral data samples. The first l samples $x_L = \{x_i\}_{i=1}^{l}$ have label $Y_L = \{y_i\}_{i=1}^{l}$, $y_i \in \{1, 2, \cdots, C\}$ and the remaining u samples $X_u = \{x_i\}_{i=l+1}^{N}$ are unlabeled. The i-th class including n_i labeled samples and labeled is C_i.

Different from traditional semi-supervised learning methods, probabilistic semi-supervised discriminant analysis [4], pays more attention to the class structure of unlabeled data. According to the assumption that samples have same structures should belong to the same category, the key idea of PSDA is to quantify the class probabilities of unlabeled samples. More specifically, in the first, it needs to

Supported by: Anhui Provincial Natural Science Foundation (Grant No.1608085MF136).

978-1-5386-6565-7/18/$31.00 ©2018 IEEE

reconstruct the unlabeled samples by an orthogonal basis matrix $B_i \in R^{Dxd}$. And orthogonal matrix $\mathbf{B}_i^D$ obtained by eigenvalue decomposition

$$(X_i - \mu_i e_{n_i}^T)(X_i - \mu_i e_{n_i}^T)^T = B_i A_i (BD)T \quad (1)$$

where μ_i is the mean value of i-th category samples, $\mu_i = (1/n_i) \sum_{X_t \in C_j} x_j$, $e_{n_i} = [1,1,\ldots,11']$. And $\mathbf{B}_i$ is formed by selecting the first d column of $\mathbf{B}_i^D$. Then, reconstructed unlabeled samples and computed reconstruction error.

$$\underline{Xi} = \mathbf{B}_i \mathbf{B}_i^T (Xi - \mu_i) + \mu_i \quad (2)$$

$$\varepsilon_i^k = X_i - \bar{X}_i \quad (3)$$

The Gauss kernel function is used to calculate class probability of each unlabeled samples belong to C category respectively.

$$w_{ik} = \exp\left(-\left(\left(\varepsilon_i^k\right)^2 / 2\sigma^2\right)\right) \quad (4)$$

According to class probabilistic matrix W, constructed weight matrix S to represent class information

$$S(i,j) = \begin{cases} I & \text{if } x_i \in X_t \text{ and } x_i \in C_j \\ W_{ij} & \text{if } X_i \in X_u \\ 0 & \text{otherwise} \end{cases} \quad (5)$$

The objective function inter-class and intra-class scatter matrix of PSDA were defined as

$$S_b = \frac{1}{N_i} \sum_{i=1}^{e} \sum_{j=1}^{N} \mathbf{s}_{ij} \left(\bar{\mu}_i - \mu\right)\left(\bar{\mu}_i - \bar{\mu}\right)^T \quad (6)$$

$$\mathbf{S}_w = \sum_{i=1}^{C} \sum_{j=1}^{N_i} \mathbf{s}_{ij} \left(x_j - \bar{\mu}_i\right)\left(x_j - \bar{\mu}_i\right)^T \quad (7)$$

where μ was global centroid of all data, and μ_i was class centroid. Finally, the transformation matrix P is obtained by solved the following maximization problem

$$J(P) = \max \left| \frac{pTS\ p}{P^T S_w^b P} \right| \quad (8)$$

III. THE PROPOSE CPSDR

CPSDR method makes full use of labeled and unlabeled data that focuses on not only class structure information of unlabeled data but also local geometry information of labeled data. First of all, the local geometry of labeled data was computed, and simultaneously, computed the class structure of unlabeled data that are used to construct a simple and effective adjacency matrix. Finally, the optimal solution was obtained by eigenvalue decomposition of the graph matrix derived by a combination of two useful information.

For labeled samples, according to two formulas of (9) and (10) calculated local information intra-class scatter S_w and inter-class scatter S_b.

$$S_W = \frac{1}{2} \sum_{i,j=1}^{n} \mathbf{W}_{i,j}^{w} (X_i - X_j)(X_i - X_j)^T \quad (9)$$

$$S_b = \frac{1}{2} \sum_{i,j=1}^{n} \mathbf{W}_{i,j}^{b} \left(\boldsymbol{x}_i - \boldsymbol{x}_{ji}\right)\left(\boldsymbol{x}_i - \boldsymbol{x}_{ji}\right)^T \quad (10)$$

where weighted matrices W'' and W^W defined by

$$W_{i,j}^{w} = \begin{cases} A_{i,j} \ln_1 & \text{if } y_i = y_j, \\ \ln & \text{if } y_i \neq y_j. \end{cases} \quad (11)$$

$$W_{i,j} = \begin{cases} A_{i,j} (\ln - \ln_1) & \text{if } y_i = y_j, \\ \ln & \text{if } Yi \neq Yj \end{cases} \quad (12)$$

among them, A is a affinity matrix, $A_{i,j}$ is the affinity value between sample x_i and sample x_j

$$\mathbf{A}_{i,j} = \exp\left(-\frac{\left\|\boldsymbol{x}_i - \boldsymbol{x}_j\right\|^2}{\sigma_i \sigma_j}\right) \quad (13)$$

The local scaling parameter σ_i defined as following, and X_1^k is the *k-th* nearest neighbor of Xi.

$$\sigma_i = \left\|\boldsymbol{x}_i - \boldsymbol{x}_i^k\right\| \quad (14)$$

For unlabeled samples, based on four formulas of (1)-(4) computed class probabilily information W, then utilized class probability to construct an adjacency matrix A as following

$$A = WW^T \quad (15)$$

where $A \in R^{UXU}$, the adjacency matrix constructed in this way can enhance the similarity of pair pixels. It is can be observed that the larger value $A_{I,J}$. of A the more possibility of two samples x_j and x_i will have the same label, otherwise it has a smaller possibility of having the same labeL At the same time, the weights of two samples were obtained.

After obtaining the adjacency matrix A, the intra-class and inter-class scatter of unlabeled data can be represented as

$$G_w = \frac{1}{2U^2} \sum_{i=1}^{U} \sum_{j=1}^{U} A_{i,j} (X_i - X_j)(X_i - X_j)^T \quad (16)$$

$$\mathbf{G}_b = \frac{1}{2u^2}\sum_{i=1}^{u}\sum_{j=1}^{u}\left(1-\mathbf{A}_{i,j}\right)\left(\boldsymbol{x}_i-\boldsymbol{x}_j\right)\left(\boldsymbol{x}_i-\boldsymbol{x}_j\right)^T \quad (17)$$

By fused the local geometry information of labeled data and the class structure information of unlabeled data, the objective function of CPSDR can be expressed by the following formula

$$p^* = \underset{\mathbf{P}\in R^{D\times}}{\text{argmax}} \frac{|pT(Sb+\lambda_1\mathbf{G}_b)p|}{|pT(s,+\lambda_2\mathbf{G}_w)p|} \quad (18)$$

where λ_1 and λ_2 are regularization parameters. And Eq.(l9) can be transformed into the following formula and it was solved by generalized eigenvalue decomposition.

$$(Sb+\lambda_1\mathbf{G}_b)\mathbf{P} = A(Sw+\lambda_2\mathbf{G}_w)\mathbf{P} \quad (19)$$

where Λ expressed eigenvalue. After solved the optional solution of Eq.(l 9), we can easily constructed the projection matrix P by the first d column of optional solution, and $d(d\ll D)$ is the dimension that should be reduced to. Finally, projection matrix P is used to reduce the dimension of original data.

$$Z = pTX \quad (20)$$

The whole procedure of CPSDR for semi-supervised HSI DR is summarized in Algorithm I.

Algorithm 1 The CPSDR Algorithm for HSI DR

Input: HSI data set $X = [x_1, x, \ldots, x_N] \in R^{DXN}$ with first l samples $\mathbf{X}_L = \{x_i\}_{i=1}^{l}$ have labels $Y_L = \{y_i\}_{i=1}^{l}$ and the remaining u samples $Xu = \{e\}_{i=l+1}^{N}$ are unlabeled.

lComputed local geometry information S wand Sb of labeled samples via Eq. (1) - (6).

2.Computed class probability W of each unlabeled samples via Eq. (8)-(11).

3.Constructed adjacency matrix A via Eq. (16).

4.Computed class structure information G_w and G_b of unlabeled samples via Eq. (17) and Eq. (18).

5.Solved generalize eigenvalue decomposition of Eq. (19), and constructed projection matrix P by first d column optimal solution, namely, $P = [P_1, P_2, \ldots, P_d] \in R^{D\times d}$.

Output: $\mathbf{Z} = pTX \in R^{dxN}$, Z is a matrix after dimensionality reduction.

IV EXPERIMENTAL RESULTS

In this section, we calculated the overall accuracy (OA) of algorithms by using the famous KNN classifier to classify the data samples in the low-dimensional space after DR, and a series of experiments based on two public HSI datasets: I) the University of Pavia data set; 2) the Salinas data set are used to evaluate the performance of the CPSDR algorithm. Especially, it is easy to find that the superiority of CPSDR compared with other semi-supervised and supervised DR algorithms, which including PCA [8-10], LDA [11, 12], LFDA [13] and SDA [14]. For CPSDR the number of nearest neighbors k is set as 7, and regularization parameters λ_1 and λ_2 are selected by cross validation. The parameters of each compared algorithm were manually adjusted to the optimum. All experiments are carried out several times independently by Monte Carlo method, and both the visual classifying results and quantitative evaluations are given for each experiment to evaluate the DR effect of different algorithms.

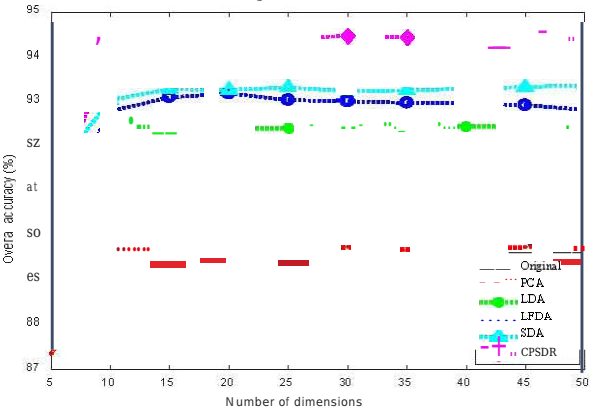

Figure. 1. OA comparison of the different DR methods on the Salinas dataset with respect to different dimensions.

A. Hyperspectral Datasets

1) Salinas dataset: The hyperspectral remote sensing images in Salinas data set are obtained by AVIRIS sensor. The image size of the data samples is 512x 217 and 224 bands. Following most studies, a total of 204 bands are used after the removal of water-absorption and noisy bands. This data-set contains 16 ground-truth classes with a total of 54129 samples.

TABLE I. OA (%) OF DIFFERENTS METHODS OF DIFFERENT LABELED TRAINING SAMPLES FOR SALINAS DATASET

Methods	Number of Labeled Training Samples				
	10	20	30	40	50
Original	84.5±2.1	88.8±O.6	90.2±O.7	*91.1±OA*	91.8±O.2
PCA	69.7±1.6	78.1±1.7	*83.6±OA*	85.9±O.5	89.5±O.5
LDA	69.6±1.8	79.2±2.2	*87A±1.6*	90.8±1.3	92.4±O.7
LFDA	61.2±3.5	88.8±1.3	91.7±O.9	92.3±O.8	92.8±O.4
SDA	85.7±1.3	89.9±O.5	91.5±O.7	92.2±O.6	93.1±O.3
CPSDR	85.5±2.3	90.1±O.9	92.7±O.3	93.5±O.6	94.2±O.3

2) PaviaU dataset: PaviaU data set is a hyperspectral remote sensing image obtained by ROSIS sensor. This data set has lIS spectral bands that range from *0.43 urn* to *0.86 urn* After removmg 12 bands that have water

absorption and noise, the remained 103 bands are used as experimental data in this study. The image size of the volumetric HIS data is 610 x 340 x 103 . Thus, there are 207400 samples with each one has a dimension of 103. Besides 164624 data samples that are corresponding to the background, the other 42776 samples belong to 9 classes of ground objects.

B. Experiments

In our experiment, 50 sample points were randomly selected from the labeled samples as the labeled training data. The overall accuracies of classification for different methods with different reduced dimensions are shown in Fig. 1, where the overall accuracy of the original data samples was used as a baseline. And the overall accuracy of each method tends to be less varied as the dimension increases and its value is around the accuracy when dimension d is equal to 10. As a comparison, the LFDA, SDA and our CPSDR methods have the best classification accuracies in all dimensional cases. Such a phenomenon could be attributed to the local geometry and class structure that both models preserved. However, our proposed CPSDR method still outperforms LFDA and SDA method in most cases, which demonstrates that the class structure is also an important property for seeking low-dimensional subspace.

In Table I and Table II, it presented some quantitative results of the HSI overall accuracies (OA) of classification experiments. We randomly selected 10, 20, 30, 40 and 50 points each category respectively for labeled training samples and the number of unlabeled training samples is fixed to be equal to the number of labeled training samples, 900 unlabeled points were used as test samples simultaneously. For a accurate comparison, all the results were reported by ten runs of independent Monte Carlo experiments when the reduced dimension d is set as 10, and it can be found that our CPSDR algorithm outperforms the other methods.

TABLE II. OA(%) OF DIFFERENT METIIODS OF DIFFERENT LABELED TRAINING SAMPLES FOR PAVIAU DATASET

Methods	Number of Labeled Training Samples				
	10	20	30	40	50
Original	70.8±0.8	73.9±0.6	*76.5±OA*	78.7±0.6	79.3±0.6
PCA	65.8±0.6	69.7±0.9	71.7±0.8	72.8±0.8	73.9±0.8
LDA	46.9±2.7	64.8±5.1	70.1±1.6	72.7±1.5	75.7±1.2
LFDA	*56.2±4A*	68.9±2.3	*72A±2.2*	73.8±0.9	*nO±1.0*
SDA	46.6±3.7	*59A±1.4*	64.7±2.9	67.8±3.1	72.0±2.7
CPSDR	65.7±2.5	76.2±1.4	79.9±1.3	82.4±0.7	83.8±0.8

V. CONCLUSION

In this paper, we proposed a novel DR method named as CPSDR for HSI. The approach learns an informative graph that well preserved not only the class structure but also local geometry of the high-dimensional HIS data set. Our method is based on LFDA algorithm, which can seek local geometry information of labeled data for discriminant analysis. To ensure the accuracy of the algorithm in the case of limited labeled samples, we utilized the class structure information of a large number of unlabeled samples. Experimental results on two public HSI data sets can verify that the proposed CPSDR is very effective in semi-supervised dimensionality reduction of HSI. Compared with the state-of-the-art methods for semi-supervised DR on HSI data, our method can provide very competitive classification performance.

ACKNOWLEDGMENTS

This work was supported by the Anhui Provincial Natural Science Foundation (Grant No.1608085MF136), National Natural Science Foundation of China (NSFC) for Youth (Grant No.61602002), and China Postdoctoral Science Foundation (Grant No.2015M582826).

REFERENCES

[1] J. Ren, J. Zabalza, S. Marshall, J. Zheng, Effective feature extraction and data reduction in remote sensing using hyperspectral imaging [applications comer], IEEE Signal Processing Magazine 31 (2014) 149-154.

[2] Z. Zhang, TW. Chow, M. Zhao, M-Isornap: Orthogonal constrained marginal isomap for nonlinear dimensionality reduction, IEEE transactions on cybernetics 43 (2013) 180-191.

[3] J.P. Cunningham, M.Y. Byron, Dimensionality reduction for large-scale neural recordings, Nature neuroscience 17 (2014) 1500-1509.

[4] W. Li, Q. Ruan, 1. Wan, Dimensionality reduction using graph-embedded probability-based semi-supervised discriminant analysis, Neurocomputing 138 (2014) 283-296.

[5] H. Huang, F. Luo, Z. Ma, H. Feng, Semi-Supervised Dimensionality Reduction of Hyperspectral Image Based on Sparse Multi-Manifold Learning, Journal of Computer & Communications 03 (2017) 33-39.

[6] P. Chen, 1. Jiao, F. Liu, 1. Zhao, Z. Zhao, S. Liu, Semi-supervised double sparse graphs based discriminant analysis for dimensionality reduction, Pattern Recognition 61 (2017) 361-378.

[7] S. Mukherjee, M. Cui, S. Prasad, Spatially Constrained Semisupervised Local Angular Discriminant Analysis for Hyperspectral Images, IEEE Journal of Selected Topics in Applied Earth Observations & Remote Sensing PP (2017) 1-10.

[8] B. Moore, Principal component analysis in linear systems: Controllability, observability, and model reduction, IEEE Transactions on Automatic Control 26 (2003) 17-32.

[9] J. Wright, A. Ganesh, S. Rao, Y. Ma, Robust Principal Component Analysis: Exact Recovery of Corrupted Low-Rank Matrices via Convex Optimization, 58 (2009) 289-298 vol.281.

[10] H. Zou, T. Hastie, R. Tibshirani, Sparse Principal Component Analysis, Journal of Computational & Graphical Statistics 15 (2006) 265-286.

[11] S.A. Awais, Face Recognition Using Principle Component Analysis(pCA) and Linear Discriminant Analysis (LDA), International Journal of Electrical & Computer Sciences (2012).

[12] J. Ye, Two-Dimensional Linear Discriminant Analysis, Advances in Neural Information Processing Systems 17 (2005) pags, 1431-1441.

[13] M. Sugiyama, T. Idé, S. Nakajima, 1. Sese, Semi-Supervised Local Fisher Discriminant Analysis for Dimensionality Reduction, Springer Berlin Heidelberg, 2008.

[14] D. Cai, X. He, 1. Han, Semi-supervised Discriminant Analysis, (2007) 1-7.