ISPRS Journal of Photogrammetry and Remote Sensing 106 (2015) 42–54

Contents lists available at ScienceDirect

ISPRS Journal of Photogrammetry and Remote Sensing

journal homepage: www.elsevier.com/locate/isprsjprs

Dimensionality reduction of hyperspectral images based on sparse discriminant manifold embedding

Hong Huang *, Fulin Luo, Jiamin Liu, Yaqiong Yang

Key Laboratory of Optoelectronic Technology and Systems of the Education Ministry of China, Chongqing University, Chongqing 400044, China

ARTICLE INFO

Article history:
Received 3 September 2014
Received in revised form 13 February 2015
Accepted 27 April 2015
Available online 20 May 2015

Keywords:
Hyperspectral image classification
Dimensionality reduction
Graph embedding
Sparse manifold learning
Discriminant learning

ABSTRACT

Sparse manifold clustering and embedding (SMCE) adaptively selects neighbor points from the same manifold and approximately spans a low-dimensional affine subspace, but it does not explicitly give a projection matrix and encounters the out-of-sample problem. To overcome this drawback, we propose a new dimensionality reduction method, called sparse manifold embedding (SME), based on graph embedding and sparse representation for hyperspectral image (HSI). It utilizes the sparse coefficients of affine subspace to construct a similarity graph and preserves this sparse similarity in embedding space. Furthermore, we try to make full use of the prior label information to design a novel supervised learning method termed sparse discriminant manifold embedding (SDME). SDME not only inherits the merits of the sparsity property of affine subspace but also boosts the compactness of intra-manifold, which achieves discriminating features and further improves the classification performance of HSI. Experiments on two real hyperspectral data sets (Indian Pines and PaviaU) show the benefits of the proposed SME and SDME methods.

1. Introduction

Hyperspectral image (HSI) is acquired by different imaging spectrometer sensors (e.g. EO-1 Hyperion, HyMap and AVIRIS), which provides detailed spectral information about objects in hundreds of spectral bands (Zhang et al., 2013). The information in HSI possesses the superior discrimination among similar ground cover classes (Chen et al., 2013). However, the large number of bands is prone to cause the curse of dimensionality, which reduces the discriminating power of ground cover as the dimensionality increases in a fixed number of training samples (Shao and Zhang, 2014). Therefore, it is an urgent issue to greatly reduce the number of bands with no appreciable loss of information.

Dimensionality reduction (DR) is an applicable way to reduce the number of bands, and it aims to find a feature mapping from the original feature space to a lower-dimensional embedding space in which some desired information can be preserved as much as possible. Some classic DR methods include principal component analysis (PCA), linear discriminative analysis (LDA) and independent component analysis (ICA). These methods are linear subspace methods, which cannot reveal the intrinsic structure of data. Recently, Bachmann and Ainsworth (2005) have discovered the intrinsic manifold in HSI data, and manifold learning methods can be used to analyze the data that lies on or near a low-dimensional manifold, such as isometric mapping (Tenenbaum et al., 2000), locally linear embedding (LLE) (Roweis and Saul, 2000) and Laplacian eigenmaps (LE) (Belkin and Niyogi, 2003). However, these methods suffer from the out-of-sample problem, which results in a difficult issue to map a new sample to the embedding space. To overcome this drawback, the nonlinear algorithms are approximately represented as the linear ones. For instance, LE and LLE are approximately linearized into locality preserving projections (LPP) (He and Niyogi, 2004) and neighborhood preserving embedding (NPE) (He et al., 2005), respectively. These methods are easy to operate for new samples, but they cannot promise good discriminating capability for their unsupervised nature.

In recent years, Yan et al. (2007) proposed a general graph embedding (GE) framework for DR. Many traditional DR algorithms (e.g. PCA, LDA, ISOMAP, LLE, LE, NPE and LPP) can be reformulated within this framework. Based on this framework, some linear discriminant methods are developed, such methods including marginal Fisher analysis (MFA) (Yan et al., 2007) and local Fisher discriminant analysis (LFDA) (Sugiyama, 2007). The differences between different GE algorithms lie in the graph

* Corresponding author.
E-mail addresses: hhuang@cqu.edu.cn (H. Huang), luofl@cqu.edu.cn (F. Luo), liujm@cqu.edu.cn (J. Liu), cquyangyq@gmail.com (Y. Yang).

http://dx.doi.org/10.1016/j.isprsjprs.2015.04.015
0924-2716/

construction and the embedded type. Therefore, an essential issue is how to compute a similarity matrix and select a constraint matrix in the framework. A popular way to construct a graph is k-nearest neighbors or ϵ-radius ball, both of which connect graph vertices with simple rules. However, it is unclear how to select the neighborhood size and define the affinity weight matrix for real-world applications.

Motivated by the recent development of sparse representation (SR), some research works have applied SR to construct a sparse similarity graph (Wright et al., 2009; Sun et al., 2013). Sparsity preserving projections (SPP) (Qiao et al., 2010) is a classic algorithm to construct a specific graph, called ℓ_1-graph, based on SR. It inherits many merits of sparse reconstruction and constructs an adapting graph for lack of model parameters, but this unsupervised nature restricts its discriminating capability. To enhance the classification performance of SPP, some supervised sparse methods are proposed, such as discriminative learning by sparse representation (DLSP) (Zang and Zhang, 2011) and sparse discriminant embedding (SDE) (Huang, 2013). However, these methods are based on the assumption that there is a single manifold in high-dimensional data set. In real applications, data sets containing multi-manifold structures are ubiquitous (Xiao et al., 2011). Elhamifar and Vidal proposed a sparse representation method in affine subspace called sparse manifold clustering and embedding (SMCE) (Elhamifar and Vidal, 2011), which can adaptively select neighbor points that lie in the same manifold and approximately span a low-dimensional affine subspace. SMCE can be applied for clustering data, but it does not directly give a projection matrix and encounters the out-of-sample problem. Furthermore, it does not explore the prior information of training samples.

In this paper, we propose a novel method termed sparse manifold embedding (SME) based on GE and SR. It constructs an adapting graph containing the similarity of data points from the same manifold, and it also overcomes the out-of-sample problem. To improve the discriminant capability of SME, we present a supervised algorithm called sparse discriminant manifold embedding (SDME), which applies the prior information to improve the separability of data from different classes. Experimental results on Indian Pines and PaviaU HSI data sets show that SME and SDME are more suitable for DR of HSI than many traditional methods.

The remainder of the paper is organized as follows. In Section 2, we review briefly GE, SR, and SMCE. SME and SDME are introduced in Section 3. Section 4 presents the experimental results to demonstrate the effectiveness of SME and SDME. Finally, we provide some concluding remarks and suggestions for future work in Section 5.

2. Related works

For convenience, we first give some notations used in this article. Let $\mathbf{X}=\{\mathbf{x}_1,\mathbf{x}_2,\ldots,\mathbf{x}_N\}$ be a set of N points in a high-dimensional space $\Re^D$ and $\ell(\mathbf{x}_i)\in\{1,2,\ldots,c\}$ denotes the class label of $\mathbf{x}_i$. The data points are assumed to lie on or near a manifold of intrinsic dimensionality d (typically $d\ll D$). DR is to find a low-dimensional embedding of $\mathbf{X}$ by mapping the D-dimensional data into a low-dimensional embedding space $\Re^d$. Let us denote the corresponding set of N points in the embedding space $\Re^d$ by $\mathbf{Y}=\{\mathbf{y}_1,\mathbf{y}_2,\ldots,\mathbf{y}_N\}$. For linear methods, the low-dimensional embedding $\mathbf{Y}$ is replaced by a matrix form $\mathbf{Y}=\mathbf{V}^T\mathbf{X}$ (where $\mathbf{V}\in\mathbf{R}^{D\times d}$).

2.1. Graph embedding

The GE framework (Yan et al., 2007) provides a unified perspective to understand most DR algorithms. In GE, it needs to construct an intrinsic graph that describes certain desired statistical or geometrical properties of data, and a penalty graph characterizes a statistical or geometric property which should be avoided. The intrinsic graph $\mathbf{G}=\{\mathbf{X},\mathbf{W}\}$ and the penalty graph $\mathbf{G}^P=\left\{\mathbf{X},\mathbf{W}^P\right\}$ are both undirected weighted graphs, where $\mathbf{X}$ is the vertex set, $\mathbf{W}\in\Re^{N\times N}$ and $\mathbf{W}^P\in\Re^{N\times N}$ are the weight matrices. The weight w_{ij} is to measure the similarity of the edge joining vertices i and j, and the weight w^P_{ij} is to suppress the similarity characteristics between the vertices i and j in the low-dimensional embedding.

The purpose of GE is to map each vertex of graph into a low-dimensional space that preserves the similarity between the vertex pairs. Then an optimal low-dimensional embedding is given by the graph preserving criterion as

$$\min_{\mathbf{y}^T\mathbf{B}\mathbf{y}=c}\frac{1}{2}\sum_{i\neq j}\|\mathbf{y}_i-\mathbf{y}_j\|^2w_{ij}=\min_{\mathbf{Y}^T\mathbf{B}\mathbf{Y}=c}\mathbf{Y}^T\mathbf{L}\mathbf{Y} \tag{1}$$

where c is a constant, $\mathbf{B}$ is the constraint matrix defined to avoid a trivial solution of the objective function, and $\mathbf{L}$ is the Laplacian matrix of graph $\mathbf{G}$. $\mathbf{B}$ typically is a diagonal matrix for scale normalization, and it may also be the Laplacian matrix of a penalty graph $\mathbf{G}^P$, that is $\mathbf{B}=\mathbf{L}^P$. The Laplacian matrix $\mathbf{L}$ and $\mathbf{L}^P$ are defined as

$$\begin{aligned}&\mathbf{L}=\mathbf{D}-\mathbf{W},\quad \mathbf{D}_{ii}=\sum_{j\neq i}w_{ij},\ \forall_i.\\&\mathbf{L}^P=\mathbf{D}^P-\mathbf{W}^P,\quad \mathbf{D}^P_{ii}=\sum_{j\neq i}w^P_{ij},\ \forall_i.\end{aligned} \tag{2}$$

2.2. Sparse representation

In recent years, SR (Wei et al., 2013; Gui et al., 2012) gives rise to the attention of many researchers, which was initially proposed as an extension to traditional signal representations such as Fourier and wavelet representations. It accounts for most or all information of a signal with a linear combination of a small number of elementary signals in an over-complete dictionary. SR has been successfully applied in signal processing, statistics and image recognition, etc.

SR has compact mathematical expression. Given a signal $\mathbf{x}_i\in\Re^D$ and a matrix $\mathbf{X}=[\mathbf{x}_1,\mathbf{x}_2,\ldots,\mathbf{x}_N]\in\Re^{D\times N}$ containing the elements of an over-complete dictionary in its columns. The purpose of SR is to represent $\mathbf{x}_i$ using as few entries of $\mathbf{X}$ as possible, and the objective function can be expressed as

$$\begin{aligned}&\min_{\mathbf{s}_i}\|\mathbf{s}_i\|_0\\&s.t.\ \|\mathbf{x}_i-\mathbf{X}\mathbf{s}_i\|<\varepsilon\end{aligned} \tag{3}$$

where ε can be seen as an error tolerance, $\mathbf{s}_i$ is the sparse coefficients of $\mathbf{x}_i$, $\|\mathbf{s}_i\|_0$ denotes the ℓ_0-norm of $\mathbf{s}_i$ which is equal to the number of non-zero components in $\mathbf{s}_i$.

However, it is a NP-hard problem to solve the Eq. (3). Researchers have discovered that if the solutions are sparse enough, and the solution of ℓ_0 minimization problem is equal to solve the ℓ_1 minimization problem as

$$\begin{aligned}&\min_{\mathbf{s}_i}\|\mathbf{s}_i\|_1\\&s.t.\ \|\mathbf{x}_i-\mathbf{X}\mathbf{s}_i\|<\varepsilon\end{aligned} \tag{4}$$

where $\|\mathbf{s}_i\|_1$ is the ℓ_1-norm of $\mathbf{s}_i$. The ℓ_1 minimization problem can be solved by LASSO or LARS (Efron et al., 2004).

2.3. Sparse manifold clustering and embedding

In this section, we introduce the SMCE algorithm. It assumes that for each data point there exists a small neighborhood in which only the points coming from the same manifold lie approximately