Neural Networks 54 (2014) 49–56

Contents lists available at ScienceDirect

Neural Networks

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

Stable locality sensitive discriminant analysis for image recognition

Quanxue Gao [a,*], Jingjing Liu [a], Kai Cui [a], Hailin Zhang [a], Xiaogang Wang [b]

[a] State Key Laboratory of Integrated Services Networks, Xidian University, Xi'an, China

[b] Department of Electronic Engineering, The Chinese University of Hong Kong, Hong Kong

ARTICLE INFO

Article history:
Received 26 July 2012
Received in revised form 27 December 2013
Accepted 21 February 2014
Available online 4 March 2014

Keywords:
Dimensionality reduction
Diversity
Similarity
Manifold learning

ABSTRACT

Locality Sensitive Discriminant Analysis (LSDA) is one of the prevalent discriminant approaches based on manifold learning for dimensionality reduction. However, LSDA ignores the intra-class variation that characterizes the diversity of data, resulting in unstableness of the intra-class geometrical structure representation and not good enough performance of the algorithm. In this paper, a novel approach is proposed, namely stable locality sensitive discriminant analysis (SLSDA), for dimensionality reduction. SLSDA constructs an adjacency graph to model the diversity of data and then integrates it in the objective function of LSDA. Experimental results in five databases show the effectiveness of the proposed approach.

1. Introduction

In many problems of computer vision, pattern recognition and machine learning, one is often confronted with high-dimensional data. However, many previous works have demonstrated that the high-dimensional data may have a lower dimensional intrinsic representation. This leads one to consider methods of dimensionality reduction that allow one to represent the data in a lower dimensional space. The aim of dimensionality reduction is to find the intrinsic geometric structure of data manifold. Thus, learning with the low dimensional manifold structure, or specifically the intrinsic topological and geometrical properties of the data manifold, becomes a crucial problem. Two of the most prevalent techniques for this purpose are Principal Component Analysis (PCA) (Jolliffe, 1986) and Linear Discriminant Analysis (LDA) (Belhumeur, Hepanha, & Kriegman, 1997; Fukunaga, 1990).

PCA is an unsupervised method. It aims to find the project directions along which the data have the maximum variance. Thus, PCA preserves the most important information that characterizes different geometric properties, namely the diversity of data. Different from PCA, LDA is supervised. It searches for the project axes on which the data points of different class are far from each other while requiring data points of the same class to be close to each other. Applied to face recognition and other high-dimensional data analysis, PCA and LDA have shown to be effective in discovering the geometric structure of image space. However, in real applications, natural images usually live in a non-line subspace, and the large distance data pairs dominate the optimal projection directions of PCA and LDA. This leads to the impairment of local geometric structure of images. Thus, they do not well discover the underlying structure of images (He, Yan, Hu, Niyogi, & Zhang, 2005; Roweis & Saul, 2000).

Recently, many geometrically motivated approaches, i.e. manifold learning approaches, have been developed for data analysis in high-dimensional spaces. Three of the most classical approaches are ISOMAP (Tenenbaum, de Silva, & Langford, 2000), Locally Linear Embedding (LLE) (Roweis & Saul, 2000), and Laplacian Eigenmaps (LE) (Belkin & Niyogi, 2003). Both LE and LLE aim to preserve the local structure of data manifold. Different from LE and LLE, ISOMAP aims to preserve the global structure of the data manifold. These approaches do yield impressive results on some benchmark artificial datasets. However, both of them yield maps that are defined only on the training data points and how to evaluate the maps on new testing points remains unclear. So, these manifold learning approaches might not be good for some computer vision and machine learning tasks, such as face recognition, document clustering (He, Yan et al., 2005; He, Yan, Hu, & Zhang, 2003). To overcome it, He et al. proposed Locality Preserving Projection (LPP) (He, Yan et al., 2005; He et al., 2003), Neighborhood Preserving Embedding (NPE) (He, Cai, Yan, & Zhang, 2005), and Isometric Projection (IsoP) (Cai, He, & Han, 2007a) by linear approximation. Both of them effectively preserve the intrinsic structure of the data and yield impressive results in face recognition, image retrieval, document clustering.

Motivated by them and LDA, many manifold learning-based discriminant approaches have been proposed. They can be roughly

* Corresponding author. Tel.: +86 13572049172.
E-mail address: qxgao@xidian.edu.cn (Q. Gao).

http://dx.doi.org/10.1016/j.neunet.2014.02.009
0893-6080/

classified into two categories: local discriminating approaches and integrating discriminating ones. The most prevalent local approaches include MFA (Margin Fisher Analysis) (Yan et al., 2007), LDE (Local Discriminant Embedding) (Chen, Chang, & Liu, 2005), LFDA (Local Fisher Discriminant analysis) (Sugiyama, 2007), LSDA (Locality Sensitive Discriminant Analysis) (Cai, He, Zhou, Han, & Bao, 2007), which was extended by imposing the uncorrelated constraint (Lu, 2010), DLA (Discriminative Locality Alignment) (Zhang, Tao, Li, & Yang, 2009), and sparse discriminant approaches (Lai, Wan, Jin, & Yang, 2011; Wang, 2012). They well preserve the local intrinsic structure and discriminate structure. Integrating approaches (Cai, He, & Han, 2007b; Huang, Xu, & Nie, 2012) obtain well-recognition accuracy by combining local and global geometric structures of data. Although their motivations are different, both of them learn the intra-class compact representation by LE (Laplacian embedding) (Roweis & Saul, 2000). LE can map nearby data points having the same class label in the original space to nearby data in the reduced space. In the ideal case, nearby data points with the same label can be mapped to a single point in the reduced space. Thus, these approaches only consider the similarity of points with the same class label and ignore the diversity, which can be learned by maximizing variance unfolding (Gao, Hao, Zhao, Shen, & Ma, 2013; Weinberger & Saul, 2006). Moreover, LE may impair the local topology of data, resulting in unstable intra-class compact representation (Gao, Liu, Zhang, Gao, & Li, 2013; Gao, Ma, Zhang, Gao, & Liu, 2013).

Many previous works have demonstrated that diversity among nearby data is also important for analyzing high-dimensional data (Gao, Hao et al., 2013; Gao, Liu, Zhang, Hou, & Yang, 2012; Hou, Zhang, Wu, & Jiao, 2009; Watanabe, Okada, & Ikeda, 2011; Weinberger & Saul, 2006). Moreover, in real applications, the intrinsic geometric structure of data is unknown and complex, and testing data is usually different from the training data due to many factors. Thus, only similarity or diversity is not sufficient to represent the intrinsic geometric structure of data. Motivated by this fact, many manifold learning-based dimensionality reduction approaches are proposed by combining similarity and diversity of data. These approaches can be roughly classified as non-discriminant (Gao, Xu, Li, & Xie, 2010; Hou et al., 2009) and discriminant approaches (Gao, Liu et al., 2013; Gao et al., 2012; Gao, Ma et al., 2013; Gao, Zhang, Yang, Liu, & Liu, 2013). However, these approaches implicitly consider that the within-class and between-class relations are equally important. This reduces the flexibility of the algorithms.

Motivated by LSDA, we propose a novel *linear* discriminant approach, namely stable locality sensitive discriminant analysis (SLSDA), which explicitly considers the similarity, diversity, and discriminating information embedded in high-dimensional data space. To be specific, we construct two adjacency graphs to model the similarity and diversity of data, respectively. We also construct the third graph to model the discriminant structure of data and then incorporate the similarity, diversity, and discriminating information into the objective function of linear dimensionality reduction. In this way, our proposed SLSDA approach well encodes the discriminating information and simultaneously obtains robust and stable intra-class compact representation, which will enhance the recognition accuracy and generalization ability of the algorithm. Extensive experiments on several image datasets indicate the effectiveness of our approach.

The rest of this paper is organized as follows. In Section 2, we provide an in-depth study for LSDA. The Stable Locality Sensitive discriminant Analysis (SLSDA) approach is introduced in Section 3. Section 4 gives the theoretical analysis of SLSDA. In Section 5, we describe some experimental results and analyses. Conclusions are summarized in Section 6.

2. LSDA

Suppose we have N training samples $x_1, x_2, \dots, x_N \in R^p$ sampled from the underlying submanifold $\boldsymbol{M}$; then the within-class graph in LSDA (Cai, He, Zhou et al., 2007) is constructed by a vertex set $X = \{x_1, x_2, \dots, x_N\}$ and weight matrix W_w. The subscription w in W_w denotes the within-class. The elements $W_{w,ij}$ in weight matrix W_w are defined as follows

$$W_{w,ij} = \begin{cases} 1, & \text{if } x_i \in N_w(x_j) \text{ or } x_j \in N_w(x_i) \\ 0, & \text{otherwise} \end{cases} \tag{1}$$

where the set $N_w(x_i)$ contains the points which are the k nearest neighbors of x_i and share the same label with x_i.

In LSDA, the within-class compactness can be learned by Laplacian embedding, i.e.,

$$\min_y J(y) = \sum_{i,j} \left(y_i - y_j\right)^2 W_{w,ij} \tag{2}$$

where $y_i = \alpha^T x_i$ is a scale and denotes the one-dimensional map of x_i. α is the projection direction.

The objective function (2) is widely used to characterize the similarity of local data. However, it results in some disadvantages in real applications.

First, it leads to unstable intrinsic structure representation. In the ideal case, the objective function (2) maps all nearby data points sharing the same class label to a single point in the reduced space. It means that Eq. (2) only considers the similarity among nearby data points and ignores the diversity of data, which is important for data classification (Gao et al., 2012, 2010; Watanabe et al., 2011; Weinberger & Saul, 2006). As previously discussed, only similarity of data is not sufficient to guarantee the stable intrinsic geometrical representation in the reduced space.

Second, it impairs the local topology of data. In real-world applications, the data distribution is uneven; thus some nearby data points may lie on a sparse region, and the distance among these data points is large. In this case, these data points with large distance will dominate the objective function (2). Thus, Eq. (2) does not guarantee that the smaller the distance between nearby data points is, the closer they should be embedded in the reduced space. This will impair the local topology of data and lead to the inexact within-class compact representation. Taking the two-dimensional data points, which are randomly selected, in Fig. 1 as an example, we plot the projection direction of Eq. (2) and one-dimensional embedding results in Fig. 1(a) and (b) respectively. We observe that Eq. (2) impairs the local topology of data in circle.

Likewise, the between-class graph is constructed by a vertex set $X = \{x_1, \dots, x_N\}$ and weight matrix W_b. The subscription b in W_b denotes the between-class. The elements $W_{b,ij}$ in weight matrix W_b can be defined as follows

$$W_{b,ij} = \begin{cases} 1, & \text{if } x_i \in N_b(x_j) \text{ or } x_j \in N_b(x_i) \\ 0, & \text{otherwise} \end{cases} \tag{3}$$

where the set $N_b(x_i)$ contains the points which are the k nearest neighbors of x_i and have different class labels.

The between-class graph is used to characterize the between-class separability by the following objective function

$$\max_y \sum_{i,j} \left(y_i - y_j\right)^2 W_{b,ij}. \tag{4}$$

The objective function (4) maps nearby data points having different class labels to be far apart in the reduced space. It is easy to see that the smaller the distance between nearby data points from different classes is, the more the discriminating information