Neurocomputing 168 (2015) 488–499

Contents lists available at ScienceDirect

Neurocomputing

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

Uncorrelated slow feature discriminant analysis using globality preserving projections for feature extraction

Xingjian Gu [a], Chuancai Liu [a,*], Sheng Wang [a], Cairong Zhao [b], Songsong Wu [c]

[a] School of Computer Science and Engineering, Nanjing University of Science and Technology, Nanjing 210094, China
[b] Department of Computer Science and Technology, Tongji University, Shanghai 201804, China
[c] School of Automation, Nanjing University of Posts and Telecommunications, Nanjing, China

ARTICLE INFO

Article history:
Received 15 December 2014
Received in revised form
6 April 2015
Accepted 20 May 2015
Communicated by Yue Gao
Available online 30 May 2015

Keywords:
Slow feature discriminant analysis
Uncorrelated feature
Face recognition
Globality preserving

ABSTRACT

Slow Feature Discriminant Analysis (SFDA) is a supervised feature extraction method for classification inspired by biological mechanism. However, SFDA only considers the local geometrical structure information of data and ignores the global geometrical structure information. Furthermore, previous works have demonstrated that uncorrelated features of minimum redundancy are effective for classification. In this paper, a novel method called uncorrelated slow feature discriminant analysis using globality preserving projections (USFDA-GP) is proposed for feature extraction and recognition. In USFDA-GP, two kinds of global information are imposed to the objective function of conventional SFDA for respecting some more global geometric structures. We also provide an analytical solution by simple eigenvalue decomposition to the optimal model instead of previous iterative method. Experimental results on Extended YaleB, CMU PIE and LFW-a face databases demonstrate the effectiveness of our proposed method.

1. Introduction

Feature extraction is a fundamental and challenging issue in pattern recognition and machine learning. It has been widely used in many practical applications [1–3], such as face recognition, image analysis and so on. The goal of feature extraction is to seek low dimensional representations of high dimensional data such that the intrinsic structures of original high dimensional data are revealed. The most famous linear feature extraction methods, such as Principal Component Analysis (PCA)[4], Linear Discriminant Analysis (LDA) [5] and their extended methods [6,7], have been extensively used for face recognition.

Linear feature extraction methods may fail to discover the nonlinear structure of data. Recently, many geometrically motivated methods have been developed to discover nonlinear structure of high dimensional data. The representative nonlinear learning methods include Isomap [8], Locally Linear Embedding (LLE) [9], and Laplacian Eigenmap (LE) [10]. Isomap preserves pairwise geodesic distance of observations in embedding space. LLE focuses on local neighborhood of each data point and preserves the minimal linear reconstructing with neighborhood in the embedding space. LE is developed on Laplace Betltrami operator to preserve proximity relationship of data points. However those manifold learning methods obtain low dimensional embedding without an explicit mapping, and they cannot extract feature beyond training samples. In order to overcome the problem, NPE [11] tries to find a linear subspace that preserves local structure based on the same principle of LLE. LPP [12] seeks a linear subspace to approximate nonlinear Laplacian Eigenmap.

In order to extract discriminant feature for classification, a lot of manifold learning based discriminant methods have been proposed to preserve the intrinsic geometry of the local neighborhoods and simultaneously reveal the discriminant structure of data [13–16]. The most prevalent approaches are Margin Fisher Analysis (MFA)[14], Discriminant Locality Preserving Projections (DLPP)[13], Local Discriminant Embedding (LDE), [15] and Local Discriminant Projection [16]. These approaches not only preserve the local geometrical structure, which represents the intra-class compactness, but also maximize the margin of inter-class to enhance the ability of classification in the reduced space.

Recently, more and more researches focus on applying the biological model to complex information tasks. Slow Feature Analysis (SFA), proposed by Wiskott and Sejnowski [17], extracted invariant from vectorial temporal signals based on temporal slowness principle. SFA has been applied for classification tasks in various ways. Franzius et al. [18] extracted the identity information of animated fish invariant to pose (including a rotation angle and the fish position) using SFA. Klampfl and Maass [19] introduced a particular Markov chain to generate a sequence of time

* Corresponding author.

http://dx.doi.org/10.1016/j.neucom.2015.05.079
0925-2312/

series used to train SFA for classification. Recently, more and more researchers have performed SFA to the applications of pattern recognition such as face recognition, human gesture recognition, human action recognition and other recognition tasks [20–25]. Zhang et al. [20] proposed the SFA framework to deal with the problem of human action recognition. SFA has a good performance on the data sets with temporal structure. However, in real applications, there are many discrete data sets that have no obvious temporal structure. In discrete scenario, it is necessary to generate time series before implementation of SFA. Huang et al. [23,25] utilized KNN criterion to construct time series and introduced Supervised Slow Feature Analysis (SSFA) for nonlinear dimensionality reduction. Gu et al. [21] proposed a new Supervised Slow feature Analysis based on Consensus Matrix (SSFACM) to construct time series for face recognition. In [22], another variant of Supervised Slow Feature that seeks the Shortest Path of each class samples (SSFASP) was proposed to construct time series for dimensionality reduction. In order to get discriminant slow feature, Huang et al. [23] propose Slow Feature Discriminant Analysis (SFDA), which minimizes within-class temporal variation and maximize between-class temporal variation simultaneously for handwritten digit recognition. The same approach is also applied more recently to human gesture recognition by Koch et al. [24]. In summary, SFDA is also a local discriminant approach. SFDA encodes the discriminative information by maximizing the distance among nearby data points from different classes and preserves the intrinsic geometrical structure by minimizing the distance among nearby data from the same class. In the ideal case, nearby points from the same class will be mapped to a single point. Thus, SFDA only captures the local structure of data and ignores the global geometrical properties, resulting in unstable intrinsic structure representation.

In real world applications, the unknown structure of data is always complex. Thus, a single local geometrical structure may not be sufficient to represent the intrinsic geometric structure of data. A reasonable approach should be one that integrates both global and local structure into the objective function of feature extraction [26–29]. LDA [5] extracts discriminant feature based on global geometric structure of data. Thus, both of LopLDA [26] and Semi-supervised Discriminant Analysis (SDA) [27] represent the local geometry by LPP [12] and then integrate the local geometry into LDA. In the literature [28], the authors proposed Joint Global and Local structure Discriminant Analysis method (JGLDA), which used two quadratic functions characterize the geometric properties of similarity and diversity of data. Zhang et al. [29] proposed Complete Global–Local LDA (CGLDA) method to incorporate three kinds of local information: local similarity information, local intra-class pattern variation and local interclass pattern variation into LDA. All of the above mentioned methods represent the global structure based on the LDA framework. However, one problem often encountered in LDA based application is that Fisher criterion is not optimal for a c-class ($c>2$) classification task. The reasons may contain two aspects. First, LDA overemphasizes the classes with larger distance in the original high dimensional space and causes large overlaps of neighboring classes in the low dimensional space [30,31]. Second, LDA assumes data covariances for all classes to be exactly identical [32], ignoring the diversity distribution of each class.

Furthermore, previous works [33–35] have demonstrated that statistically correlated features contain redundancy, which may distort the distribution of the feature and even dramatically degrade the performance. Recently, several uncorrelated discriminant methods have been developed. Jin et al. [33] proposed an uncorrelated linear discrimination analysis (ULDA) approach which maximizes Fisher criterion and simultaneously produces statistical uncorrelated features. To explore local information, Jing et al.[34] proposed a feature extraction approach named Local Uncorrelated Discriminant Transform (LUDT) for face recognition by constructing the local uncorrelated constraints and calculating the optimal discriminant vectors. However, the above mentioned uncorrelated methods are all implemented in an iterative way and it needs to take a long time to complete the iterative process.

In this paper, we propose a novel feature extraction method namely Uncorrelated Slow Feature Discriminant Analysis using Globality Preserving Projection (USFDA-GP), which integrates globality information into SFDA and extracts statically uncorrelated discriminant feature for classification. Some aspects of the proposed USFDA-GP method are worth highlighting.

1. This paper proposes a novel uncorrelated slow feature discriminant analysis method. The proposed method integrates globality information into traditional SFDA and removes the redundancy between features to enhance the discriminant ability.
2. USFDA-GP offers analytical solutions using standard eigenvalue decomposition and avoids the iterative process which is very time consuming. That is, USFDA-GP is more efficient in computation.
3. The features extracted by USFDA-GP, which integrates the global information of data into SFDA, are proven to have good discrimination ability. A series of experimental results show that USFDA-GP has many advantages on efficiency and accuracy in classification tasks.

The rest of the paper is organized as follows. In Section 2, we briefly review LDA and SFDA. In Section 3, we give the motivations of Uncorrelated Slow Feature Discriminant Analysis using Globality Preserving Projection and describe it in detail. In Section 4 , experiments with face image databases are carried out to demonstrate the effectiveness of the proposed method. Finally, the conclusions are made in Section 5.

2. Related work

Given a sample set $X=\{x_1,x_2,\ldots,x_n\}\in R^{D\times n}$ and samples belong to one of c classes $\{X_1,X_2,\ldots,X_c\}$. Let c denote the total number of classes and n_i denote the number of training samples in the ith class. Let x_i^j denote the jth sample in the ith class, $\overline{x}$ denote the mean of all training samples, $\overline{x}_i$ be the mean of the ith class.

2.1. LDA

LDA extracts discriminant feature for classification based on the principle of maximizing between-class scatter and minimizing within-class scatter simultaneous. The between-class and within-class scatter matrices can be evaluated as follows:

$$S_b=\frac{1}{c}\sum_{i=1}^{c}(\overline{x}_i-\overline{x})(\overline{x}_i-\overline{x})^T \tag{1}$$

$$S_w=\frac{1}{n}\sum_{i=1}^{c}\sum_{j=1}^{n_i}(x_i^j-\overline{x}_i)(x_i^j-\overline{x}_i)^T \tag{2}$$

The LDA based discriminant rule is defined as follows:

$$W^*=\arg\max_W\frac{\mathrm{tr}(W^TS_bW)}{\mathrm{tr}(W^TS_wW)} \tag{3}$$

The solution to Eq. (3) can be solved by generalized eigenvalue problem $S_bw=\lambda S_ww$ and optimal projections can be selected as eigenvectors $w_1,w_2,\ldots,w_d$ corresponding to the first largest eigenvalues $\lambda_1,\lambda_2,\ldots,\lambda_d$.