Neural Networks 55 (2014) 83–97

Contents lists available at ScienceDirect

Neural Networks

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

A general soft label based Linear Discriminant Analysis for semi-supervised dimensionality reduction

Mingbo Zhao [a], Zhao Zhang [b,*], Tommy W.S. Chow [a], Bing Li [c]

[a] Department of Electronic Engineering, City University of Hong Kong, Tat Chee Avenue, Kowloon, Hong Kong Special Administrative Region
[b] School of Compute Science and Technology, Soochow University, Suzhou 215006, China
[c] School of Economics, Wuhan University of Technology, Wuhan 430070, China

HIGHLIGHTS

- A label propagation procedure is proposed to handle the outliers and dataset with multi-density distribution.
- A soft label based LDA is proposed to handle the out-of-sample problem of label propagation.
- A fast solution for solving SL-LDA is presented based on a weighted and regularized least square.
- A flexible version of SL-LDA is proposed for better cope with the data sampled from a nonlinear manifold.
- Simulation results show the efficiency and effectiveness of the proposed methods.

ARTICLE INFO

Article history:
Received 26 April 2013
Received in revised form 1 March 2014
Accepted 13 March 2014
Available online 13 April 2014

Keywords:
Linear Discriminant Analysis
Semi-supervised dimension reduction
Soft label
Label propagation

ABSTRACT

Dealing with high-dimensional data has always been a major problem in research of pattern recognition and machine learning, and Linear Discriminant Analysis (LDA) is one of the most popular methods for dimension reduction. However, it only uses labeled samples while neglecting unlabeled samples, which are abundant and can be easily obtained in the real world. In this paper, we propose a new dimension reduction method, called "SL-LDA", by using unlabeled samples to enhance the performance of LDA. The new method first propagates label information from the labeled set to the unlabeled set via a label propagation process, where the predicted labels of unlabeled samples, called "soft labels", can be obtained. It then incorporates the soft labels into the construction of scatter matrixes to find a transformed matrix for dimension reduction. In this way, the proposed method can preserve more discriminative information, which is preferable when solving the classification problem. We further propose an efficient approach for solving SL-LDA under a least squares framework, and a flexible method of SL-LDA (FSL-LDA) to better cope with datasets sampled from a nonlinear manifold. Extensive simulations are carried out on several datasets, and the results show the effectiveness of the proposed method.

© 2014 Elsevier Ltd. All rights reserved.

1. Introduction

Dealing with high-dimensional data is always a major problem in the research of pattern recognition and machine learning. Typical applications of these include face recognition, document categorization, and image retrieval. Finding a low-dimensional representation of high-dimensional space, i.e., dimension reduction, is thus of great practical importance. The goal of dimension reduction is to reduce the complexity of input space and embed high-dimensional space into a low-dimensional space, while retaining most of the desired intrinsic information (Huang, Xu, & Nie, 2012; Jia, Nie, & Zhang, 2009; Nie, Xiang, Jia, & Zhang, 2009; Roweis & Saul, 2000; Tenenbaum, Silva, & Langford, 2000; Yan et al., 2007; Zhang, Chow, & Zhao, 2013; Zhang, Tao, Li, & Yang, 2009; Zhang, Yan, & Zhao, 2014; Zhang, Zhao, & Chow, 2012, 2013; Zhao, Zhang, & Chow, 2012; Zhao, Zhang, Chow, & Li, 2014). Among all the dimension reduction techniques, Principle Component Analysis (PCA) (Turk & Pentland, 1991) and Linear Discriminant Analysis (Belhumeur, Hespanha, & Kriegman, 1997) are the most popular methods and have been widely used in many classification applications. The objective of PCA is to pursue a set of orthogonal bases that can capture the directions of maximum variance in the dataset for optimal reconstruction. On the other hand, LDA uses the

* Corresponding author. Tel.: +86 852 3442 2874.
E-mail addresses: mbzhao4@gmail.com (M. Zhao), cszzhang@gmail.com (Z. Zhang), eetchow@cityu.edu.hk (T.W.S. Chow), lib675@163.com (B. Li).

http://dx.doi.org/10.1016/j.neunet.2014.03.005
0893-6080/© 2014 Elsevier Ltd. All rights reserved.

within-class scatter matrix S_w to evaluate the compactness within each class and between-class scatter matrix S_b to evaluate the separation between different classes. The objective of LDA is then to find the optimal projection that maximizes the between-class scatter matrix S_b while minimizing the within-class scatter matrix S_w in the low-dimensional subspace. Given that the within-class scatter matrix is nonsingular, the optimization problem of LDA can be solved by generalized eigenvalue decomposition (GEVD), i.e., finding the d largest eigenvectors corresponding to the eigenvalues of $S_w^{-1}S_b$ (Fukunaga, 1990). However, the within-class scatter matrix that characterizes the within-class compactness is not of full rank; thus, it is not invertible and the optimal projection matrix may be found to be incorrect. Many variants of LDA have been proposed to solve this, including Regularized LDA (Friedman, 1989), Null-space LDA (Chen, Liao, Ko, Lin, & Yu, 2000), Uncorrelated LDA (Ye, Janardan, Li, & Park, 2006), and SRDA (Cai, He, & Han, 2008).

To handle the nonlinear classification problem, another type of method maps the data samples in Euclidean space to a higher-dimensional Reproducing Kernel Hilbert Space (RKHS), and then finds a subspace in RKHS using the Representer theorems. Typical methods include Kernel PCA (KPCA) and Support Vector Machine (SVM). The KPCA, which is an extension of PCA, is one of the earliest kernel methods (Scholkopf, Smola, & Muller, 1998). Being a supervised method, SVM (Cortes & Vapnik, 1995; Vapnik, 1995) is based on structural risk minimization principles (SRM) (Vapnik, 1998), which strike a balance between traditional empirical risk and model complexity. This is implemented by finding an optimal separating hyperplane, giving a maximum margin between the positive and negative training samples that are closest to the hyperplane. Furthermore, to handle the nonlinear problem, the input space is first mapped into RKHS, wherein a linear SVM is constructed. Due to the above merits, SVM shows great performance in various real-world applications, such as text categorization, face recognition, and image retrieval (Li, Allinson, Tao, & Li, 2006; Zhang, Lin, & Zhang, 2001).

In general, although supervised methods outperform unsupervised methods due to the utilization of label information, they need a large number of labeled samples to produce satisfactory results (Belhumeur et al., 1997). However, in many cases, obtaining enough labeled samples for training can be problematic because labeling samples is time-consuming and costly. However, unlabeled samples are abundant and can be easily obtained in the real world. Thus, semi-supervised learning methods (Belkin, Niyogi, & Sindhwani, 2006; Cai, He, & Han, 2007a, 2007b; Chen, Ye, & Li, 2007; Nie, Xu, Tsang, & Zhang, 2010; Wang, Wang, Zhang, Shen, & Quan, 2009; Wang & Zhang, 2008; Zhou, Bousquet, Lal, Weston, & Scholkopf, 2004; Zhu, Ghahramani, & Lafferty, 2003), which incorporate both labeled and unlabeled samples into learning procedures, have become more effective than only relying on supervised learning. Many semi-supervised methods have been proposed in the past few years. These methods include Gaussian Fields and Harmonic Functions (GFHF) (Zhu et al., 2003), Learning with Local and Global Consistency (LLGC) (Zhou et al., 2004), General Graph-based Semi-supervised Learning (Nie, Xiang, Liu, & Zhang, 2010), Semi-supervised Discriminant Analysis (SDA) (Cai et al., 2007a), Laplacian Linear Discriminant Analysis (Chen et al., 2007), and Laplacian Least Square (LapRLS) (Belkin et al., 2006).

The above methods can be divided into two categories: transductive methods and inductive methods. Two well-known transductive methods are GFHF and LLGC. These algorithms work in a transductive way by propagating label information from labeled set to unlabeled set. However, they cannot predict the class labels of new-coming samples hence result in the out-of-sample problem. In contrast, inductive methods, e.g., SDA, LapRLS, and LapLDA can solve this problem. These algorithms will firstly construct a manifold regularized term to preserve the geometrical structure with both labeled and unlabeled set (He, Yan, Hu, Niyogi, & Zhang, 2005). Then, they will find a transformed matrix to perform dimension reduction by incorporating the manifold regularized term into the original objective function of supervised algorithms. Hence, the new-coming samples can be projected into a low-dimensional subspace by using such a transformed matrix and the out-of-sample problem can be solved, which is more practical in real-world applications.

In this paper, we propose a new semi-supervised dimension reduction method, called "SL-LDA", to enhance conventional LDA performance by incorporating soft labels into scatter matrixes. The proposed method first propagates label information from the labeled set to unlabeled set via label propagation, where the predicted class labels of unlabeled samples, called "soft labels", can be obtained. It then finds a transformed matrix to perform dimension reduction by incorporating the soft labels into the scatter matrixes. The proposed method can be viewed as a unified framework to extend many existing LDA algorithms to their semi-supervised versions. Its basic ideas are different from those semi-supervised algorithms, such as SDA, LapLDA, and LapRLS. These methods use the labeled and unlabeled samples in a simple manner to construct a manifold regularized term and add it to the objective function of supervised algorithms. However, in the proposed algorithm, by incorporating the soft labels into training, it can well preserve the probability distribution of samples both in labeled and unlabeled sets; thus, a better subspace for dimension reduction is obtained. We also analyze our proposed algorithm under a least squares framework. It can be concluded that given a certain class indicator, the optimization problem of the proposed method can be equivalent to a weighted least squares problem.

In addition, we observe that in SL-LDA, the low-dimensional data representation Z is constrained to be in the linear space spanned by the training data matrix, which is over-strict to fit the data sampled from a nonlinear manifold (Nie, Xu et al., 2010). To handle this problem, we relax this hard constraint by introducing a regression residual term into the reformulated objective function of SL-LDA to enforce the low-dimensional data representation Z to be close to the projected training data after using the projection matrix. With such relaxation, a flexible method of SL-LDA, referred to as "FSL-LDA", is proposed and generally suitable to cope with a certain type of nonlinear manifold that is somewhat close to a linear subspace.

The main contributions of this paper are summarized as follows:

(1) We present an effective label propagation procedure, which is based on a new local reconstruction graph with symmetrization and normalization. The symmetrization can add more connections of a sample with its neighborhoods, and the normalization can handle the case when the density of the dataset varies dramatically.
(2) The proposed SL-LDA can preserve more discriminative information embedded into the soft labels of unlabeled samples than other methods, which is advantageous for the performance of classification. The proposed method can also be easily extended to its kernel version by using kernel tricks (Yang, Frangi, Yang, Zhang, & Jin, 2005; Zhang, Nie, & Xiang, 2010).
(3) Motivated by the equivalence between LDA and the least squares problem (Sun, Ceran, & Ye, 2010; Ye, 2007; Zhang, Dai, Xu, & Jordan, 2010), we extend this relationship and further analyze SL-LDA under a weighted least squares problem (W-LS). Based on such relationship, we then propose a more efficient approach for calculating the optimal solution of SL-LDA, which is a linear system of equation and can be performed on large-scale datasets.

Download English Version:

https://daneshyari.com/en/article/403962

Download Persian Version:

https://daneshyari.com/article/403962

Daneshyari.com