Neurocomputing 135 (2014) 284–298

Contents lists available at ScienceDirect

Neurocomputing

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

An approximate closed-form solution to correlation similarity discriminant analysis

Taiping Zhang [a,*], Yuan Yan Tang [b], C.L. Philip Chen [b], Zhaowei Shang [a], Bin Fang [a]

[a] College of Computer Science, Chongqing University, Chongqing 400030, China
[b] Faculty of Science and Technology, University of Macau, Macau, China

ARTICLE INFO

Article history:
Received 1 April 2013
Received in revised form
9 December 2013
Accepted 17 December 2013
Communicated by S. Fiori
Available online 31 January 2014

Keywords:
Correlation
Similarity measure
Similarity discriminant analysis
Linear discriminant analysis
Feature extraction

ABSTRACT

High-dimensional data often lie on relatively low-dimensional manifold, while the nonlinear geometry of that manifold is often embedded in the similarities between the data points. Correlation as a similarity measure is able to capture these similarity structures. In this paper, we present a new correlation-based similarity discriminant analysis (CSDA) method for class separability problem. Firstly, a new formula based on the trace of matrix is proposed for computing the correlation between data points. Then a criterion *maximizing the difference between within-class correlation and between-class correlation* is proposed to achieve maximum class separability. The optimization of the criterion function can be transformed to an eigen-problem and an approximate closed-form solution can be obtained. Theoretical analysis shows that CSDA can be interpreted in the framework of graph-based learning. Furthermore, we also show how to extend CSDA to a nonlinear case through kernel-based mapping. Extensive experiments on different data sets are reported to illustrate the effectiveness of the proposed method in comparison with other competing methods.

1. Introduction

Linear discriminant analysis (LDA) is a well-known method for feature extraction and dimensionality reduction. It can extract a low-dimensional feature representation of the data by minimizing the within-class variance and maximizing the between-class variance simultaneously. LDA has been widely applied to many modern applications in the field of pattern recognition and computer vision, such as face recognition [1–3] and text/document classification [4,5].

Euclidean distance is a metric which measures the dissimilarity, rather than the similarity, between data points. Since LDA uses Euclidean distance as a distance measure, it extracts the discriminant information hidden in widely separated data points, instead of that hidden in nearby data points. As a result, LDA cannot deal with the merging classes which are close together and far away from other classes [6], as illustrated in Fig. 1(b). Hence, classical LDA does not always find the optimal projection direction for pattern classification [7]. In order to improve the performance of LDA, Hamsici and Martinez [7] provided a algorithm to find that d-dimensional discriminant subspace, where the Bayes error is minimized for the C class problem with homoscedastic Gaussian distributions. Lotlikar and Kothari [8] developed a fractional-step LDA (FS-LDA) by introducing a weighting function. This is equivalent to replacing the Euclidean distance in the LDA criterion with a weighted Euclidean distance. In [9], Loog et al. [9] used the Mahanalobis distance to replace the Euclidean distance in the LDA criterion, which is called the approximate pairwise accuracy criterion (aPAC). While FS-LDA and aPAC attempt to overcome the drawback of Euclidean distance by using weighting functions, it is difficult to determine what the best weighting function should be.

It is known that there is a low-dimensional nonlinear manifold lying in the high-dimensional data [10–13], while the dimensionality and nonlinear geometry of that manifold is often embedded in the similarities between the data points [13]. An effective manifold learning method must be able to find a low-dimensional representation of the data that best preserves the similarities between data points and can effectively capture the intrinsic geometrical structure of the data. In order to capture the latent manifold structure of data, many feature extraction and dimensionality reduction methods based on other distance measures were proposed in the past decade. The isometric mapping (Isomap) method [14] substitutes geodesics distance for Euclidean distance and then classical multidimensional scaling (MDS) is used to find the optimum low-dimensional representation. The Laplacian eigenmap (LE) [15] and locality preserving projection (LPP) [16] methods try to solve a constrained optimization problem in order to minimize a weighted sum of squared local distances between neighbor data points. The distance measure reflects the average of the Laplacian Beltrami operator on tangent spaces over the manifold which can be considered as a weighted Euclidean distance.

In an effective feature extraction method, a proper measure is the key to capturing the latent intrinsic structures. In order to

* Corresponding author. Tel.: +86 23 65112784; fax: +86 23 65102502.
E-mail address: tpzhang@cqu.edu.cn (T. Zhang).

0925-2312/$ - see front matter © 2014 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.neucom.2013.12.015

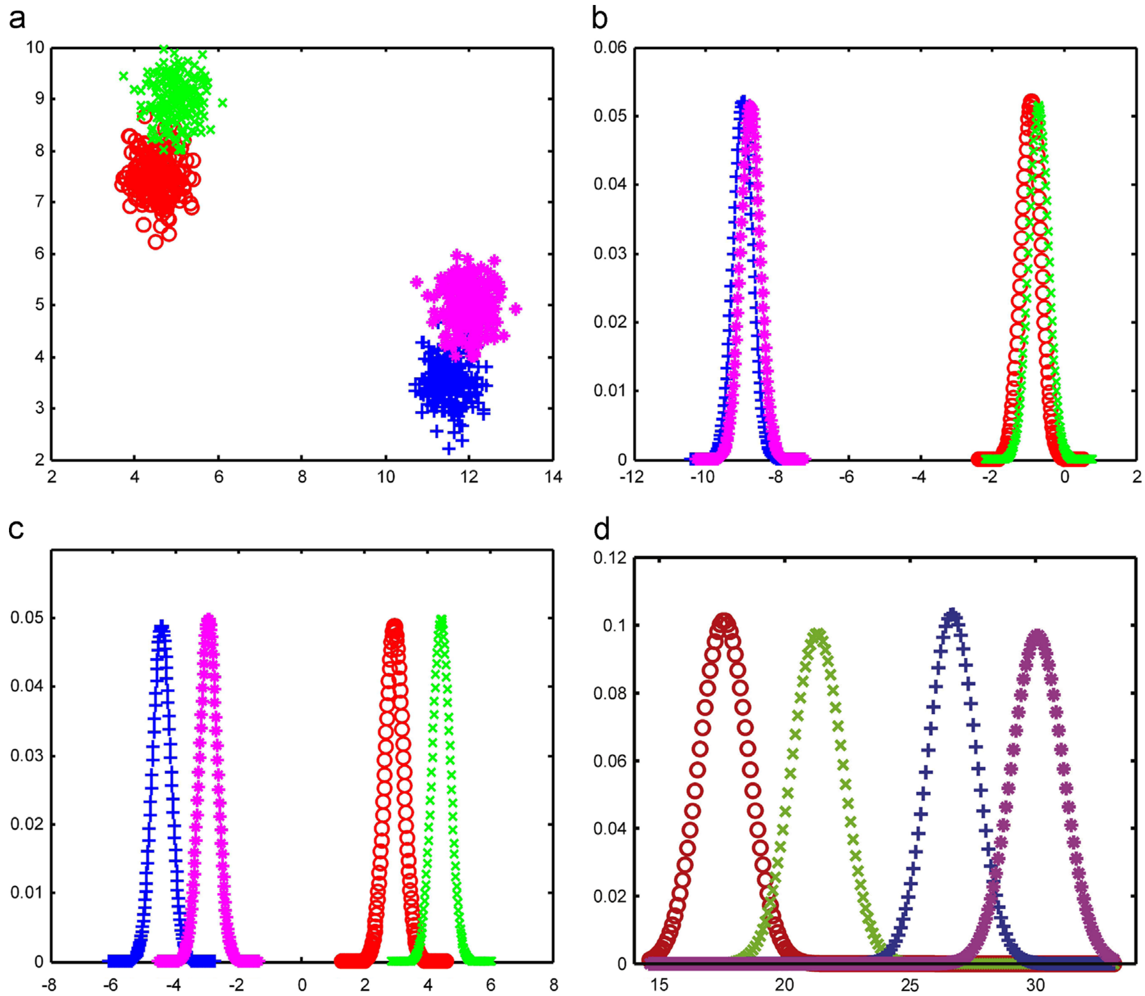

Fig. 1. (a) shows four data classes drawn from the Gaussian distribution in $\mathbb{R}^2$, denoted by "x", "o", "$*$", "$+$". (b) shows the LDA projection result in 1D subspace. (c) shows the 1D projection obtained by our proposed CSDA. (d) shows the 1D projection obtained by BLDA.

effectively capture the intrinsic manifold structure embedded in the similarities between data points, correlation as a similarity measure could be a suitable measure. In probability theory and statistics, correlation indicates the strength and direction of a linear relationship between two random variables which reveals the nature of data represented by the classical geometric concept of an "angle". It is a scale-invariant association measure usually used to calculate the similarity between two vectors. The correlation between two vectors (column vectors) $\mathbf{u}$ and $\mathbf{v}$ is defined as

$$Corr(\mathbf{u},\mathbf{v}) = \frac{\mathbf{u}^T\mathbf{v}}{\sqrt{\mathbf{u}^T\mathbf{u}}\sqrt{\mathbf{v}^T\mathbf{v}}} = \left\langle \frac{\mathbf{u}}{\|\mathbf{u}\|}, \frac{\mathbf{v}}{\|\mathbf{v}\|} \right\rangle, \tag{1}$$

where the operator $\langle \cdot, \cdot \rangle$ means the inner product of two vectors. The correlation corresponds to an angle θ such that $\cos\theta = Corr(\mathbf{u},\mathbf{v})$. The larger the value of $Corr(\mathbf{u},\mathbf{v})$, the stronger is the association between the two vectors $\mathbf{u}$ and $\mathbf{v}$. In recent years, some correlation-based feature extraction and dimensionality reduction methods have been reported in the literature [17–21]. It should be pointed out that the correlation discriminant analysis (CDA) [17] and the correlation embedding analysis (CEA) [18,22] are two discriminant analysis methods in the correlation measure space. However, they employ the gradient descent technique to find a local optimum solution. Since the objective function is not convex, the gradient descent technique cannot guarantee to get the desired result. Usually, many runs are performed with random initial values and the best result is selected. This results in long processing time and high computation cost. In addition, the selection of step size is a difficult task. Consequently, the CDA and CEA methods have limited usage in practice.

In this paper, we present a new method for correlation-based similarity discriminant analysis (CSDA). We first propose a new correlation calculation formula to compute the within-class correlation and the between-class correlation. Then the maximization of the difference between within-class correlation and between-class correlation is studied and a criterion is proposed. We show that the optimization of the proposed criterion can be transformed to an eigen-problem, from which an approximate closed-form solution can be derived. Different from Euclidean distance, correlation is a similarity measure, CSDA can effectively capture the intrinsic manifold structure embedded in the similarities between data points. As shown in Fig. 1(c), CSDA can find that 1D subspace where the data is more separated than that obtained by classical LDA based on Euclidean distance. Theoretical analysis shows that CSDA can be interpreted in the framework of graph-based learning. Moreover, while it is difficult to extend the CDA and CEA methods to kernel versions, the kernel variants of CSDA can be easily derived.

The remainder of this paper is organized as follows. In Section 2, some related works are briefly reviewed. The CSDA method is proposed in Section 3, followed by a theoretical analysis in Section 4. The kernel extension of CSDA is presented in Section 5. In Section 6, experimental results are provided to illustrate the performance of our method by comparing it with existing methods. Finally, conclusions are drawn in Section 7.