Neurocomputing 151 (2015) 1255–1261

Contents lists available at ScienceDirect

Neurocomputing

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

MiLDA: A graph embedding approach to multi-view face recognition

Yiwen Guo [a,b,*], Xiaoqing Ding [a], Jing-Hao Xue [b]

[a] State Key Laboratory of Intelligent Technology and Systems, Tsinghua National Laboratory for Information Science and Technology, Department of Electronic Engineering, Tsinghua University, Beijing 100084, PR China

[b] Department of Statistical Science, University College London, London WC1E 6BT, UK

ARTICLE INFO

Article history:
Received 19 August 2014
Received in revised form
28 October 2014
Accepted 2 November 2014
Communicated by X. Gao
Available online 10 November 2014

Keywords:
Face recognition
Graph embedding
Multi-view

ABSTRACT

In a vast number of real-world face recognition applications, gallery and probe image sets are captured from different scenarios. For such multi-view data, face recognition systems often perform poorly. To tackle this problem, in this paper we propose a graph embedding framework, which can project the multi-view data into a common subspace of higher discriminability between classes. This framework can be readily utilized to extend classical dimensionality reduction methods to multi-view scenarios. Hence, by utilizing the framework for multi-view face recognition, we propose multi-view linear discriminant analysis (MiLDA). We also empirically demonstrate that, for several distinct multi-view face recognition scenarios, MiLDA has an excellent performance and outperforms many popular approaches.

© 2014 Elsevier B.V. All rights reserved.

In a vast number of real-world face recognition applications, gallery and probe image sets are often captured from different scenarios. This results in large intra-class variations across the gallery and probe image sets. For instance, face images in the gallery set are often full-faced images with high quality, while those in the probe set are often of lower resolution and with pose and illumination variations. Hence, recognition of faces in these scenarios, sometimes called "multi-view face recognition", remains a big challenge, in spite of the fact that over the past decades tremendous progress has been made for face recognition under controlled environment. This challenge usually makes traditional face recognition systems suboptimal or even out of operation, and thus is receiving more and more attention [1–13]. Other multi-view face recognition scenarios include still-to-video face recognition [10] and heterogeneous face recognition (e.g., sketch-to-photo [14–20] and near infrared vs. visible light [20–22]).

Technically, image samples from different views usually span different feature spaces. This makes cross-view sample comparison, and thus multi-view recognition, extremely difficult. To tackle this problem, two main strategies are primarily designed: one is through extraction of cross-view invariant features [9,18] and the other is through data transformation [2–4,7,10–12,14,16,23,24,8,13]. However, cross-view invariant features may not be easily extracted and their discriminability may not be guaranteed high. Hence, data transformation is often preferred as it is more efficient and readily interpretable.

Among various data transformation methods, the subspace approaches aiming at learning view-specific projection directions are becoming more and more popular [2,11,12,8,13]. These approaches use samples from different views to learn the projection directions, such that a common discriminative subspace can be formed and accurate recognition can be achieved in the new subspace.

In this paper, we propose a novel subspace learning framework for multi-view data analysis on the basis of graph embedding [25]. More specifically, we introduce a new measure of distance between projected vertex sets of intrinsic graphs, to mitigate the effect of differences between views and preserve the intrinsic graphs. The objective of the new framework, as shown in Fig. 1, is to find a common subspace, which not only preserves certain statistical and geometric properties of the original data but also becomes more discriminative.

The proposed framework is straightforward and effective to extend various classical dimensionality reduction methods to multi-view scenarios. Moreover, it can be efficiently solved using a greedy algorithm and the constrained concave–convex procedure taken for each pair of projection directions. Hence, by utilizing the framework for multi-view face recognition, we propose multi-view linear discriminant analysis (MiLDA) and, compared with some popular approaches, MiLDA demonstrates an excellent performance.

The rest of this paper are structured as follows. We introduce some popular subspace approaches for multi-view data analysis in Section 1. We then present in Section 2 our graph embedding framework, its algorithmic solution, and its specific approach proposed for multi-view face recognition (i.e. MiLDA). Section 3

* Corresponding author at: State Key Laboratory of Intelligent Technology and Systems, Tsinghua National Laboratory for Information Science and Technology, Department of Electronic Engineering, Tsinghua University, Beijing 100084, PR China. Fax: +86 10 62772044 211.

E-mail addresses: guoyiwen@ocrserv.ee.tsinghua.edu.cn (Y. Guo), dxq@ocrserv.ee.tsinghua.edu.cn (X. Ding), jinghao.xue@ucl.ac.uk (J.-H. Xue).

http://dx.doi.org/10.1016/j.neucom.2014.11.004
0925-2312/© 2014 Elsevier B.V. All rights reserved.

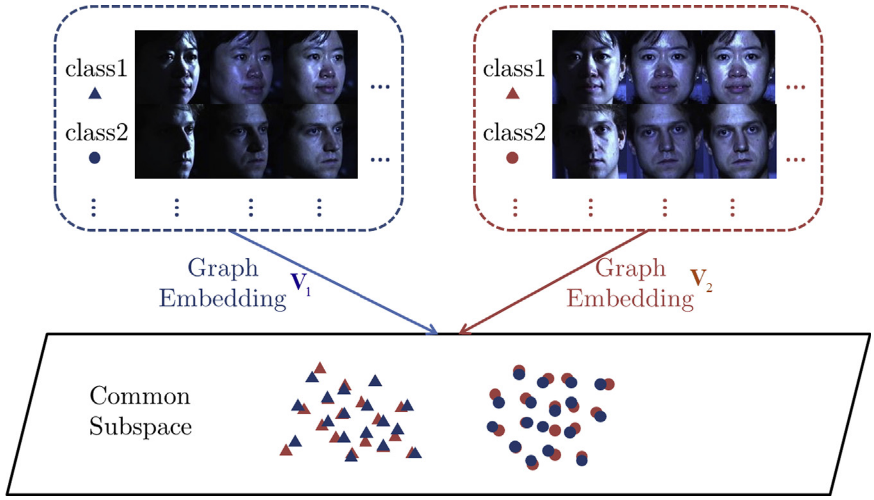

Fig. 1. Illustration of the proposed graph embedding multi-view approach to face recognition. Face images in the blue rounded rectangle are in 45° pose, while those in the red rounded rectangle are frontal face images. $\mathbf{V}_1$ and $\mathbf{V}_2$ are two projection directions learned by our algorithm. (For interpretation of the references to color in this figure caption, the reader is referred to the web version of this paper.)

experimentally evaluates MiLDA for two multi-view face recognition scenarios and Section 4 draws some conclusions.

1. Related works

1.1. Canonical correlation analysis

The most famous subspace-based multi-view data analysis approach is canonical correlation analysis (CCA) [26], the goal of which is to extract pairs of directions that maximize correlation between two data sets.

Formally, given two matrices $X_1 = [x_1^{(1)}, x_1^{(2)}, \ldots, x_1^{(N)}]$ in which $x_1^{(i)} \in \mathbb{R}^{d_1}$ and $X_2 = [x_2^{(1)}, x_2^{(2)}, \ldots, x_2^{(N)}]$ in which $x_2^{(i)} \in \mathbb{R}^{d_2}$, CCA iteratively learns the pairs of normalized projection directions uncorrelated with the pairs previously learned. The first pair of projection directions v_1 and v_2 can be found through

$$\max_{v_1, v_2} \frac{v_1^T X_1 X_2^T v_2}{\sqrt{v_1^T X_1 X_1^T v_1}\sqrt{v_2^T X_2 X_2^T v_2}} \quad \text{s.t.} \quad v_1^T v_1 = 1, \quad v_2^T v_2 = 1, \tag{1}$$

which can be solved via eigenvalue decomposition.

CCA can transform the original face images to a common subspace and thus make the cross-view comparison possible, for example, for face recognition across pose differences [2].

1.2. Partial least squares and bilinear models

The subspace learned by CCA may not be optimal in many cases. Other popular approaches to the learning of an optimal subspace include partial least squares (PLS) and bilinear models (BLM).

To some extent, PLS is similar to CCA if the variance within views are not influential. The main difference between PLS and CCA is that PLS seeks the projection directions that maximize the covariance between data sets while CCA goes for the correlation. That is to say, the first pair of projection directions of PLS can be found through

$$\max_{v_1, v_2} v_1^T X_1 X_2^T v_2 \quad \text{s.t.} \quad v_1^T v_1 = 1, \quad v_2^T v_2 = 1. \tag{2}$$

Unlike CCA and PLS, BLM uses singular value decomposition (SVD) of matrix YY^T, in which $Y = [X_1^T, X_2^T]^T$ is the concatenation matrix of X_1 and X_2, to derive a common subspace for multi-view data. BLM can also be written as an iterative procedure and the first pair of projection directions v_1 and v_2 can be found through

$$\max_{v_1, v_2} v_1^T X_1 X_1^T v_1 + v_2^T X_2 X_2^T v_2 + 2 v_1^T X_1 X_2^T v_2 \quad \text{s.t.} \quad v_1^T v_1 + v_2^T v_2 = 1. \tag{3}$$

In [11], both BLM and PLS are found to achieve good performance for several multi-view face recognition tasks.

1.3. Generalized multiview analysis

CCA, PLS and BLM are all unsupervised approaches in our context as they do not take class labels into account. This makes their performance suboptimal for some complex multi-view recognition tasks. Recently, several supervised multi-view analysis approaches inspired by the classical discriminant analysis methods have been proposed [27,28]. However, these approaches use multi-view data mainly to learn better classifiers for single-view data. Furthermore, most of these approaches cannot be generalized to classification of data from unseen classes, which restricts the use of them. Considering these drawbacks, a generalized multi-view analysis (GMA) model is proposed [12] as a supervised extension of CCA.

To apply GMA to multi-view face recognition, the authors of [12] further propose the generalized multiview LDA (GMLDA)

$$\max_{v_1, v_2} \mu v_1^T X_1 W_1 X_1^T v_1 + \gamma v_2^T X_2 W_2 X_2^T v_2 + v_1^T X_1 X_2^T v_2 \quad \text{s.t.} \quad v_1^T P_1 v_1 + \eta v_2^T P_2 v_2 = 1, \tag{4}$$

in which W_l and P_l are $N_l \times N_l$ matrices for $l = 1,2$. Let N_l^k be the number of kth class samples from the lth view and I_l be the $N_l \times N_l$ identity matrix, then each element of the matrices W_l and P_l can be calculated as

$$W_l^{(ij)} = \begin{cases} 1/N_l^k & \text{if } x_l^{(i)} \text{ and } x_l^{(j)} \text{ both belong to class } k \\ 0 & \text{otherwise,} \end{cases} \tag{5}$$

$$P_l = I_l - W_l. \tag{6}$$

The optimization problem (4) can be rewritten by introducing a Lagrange multiplier and solved via generalized eigenvalue decomposition. The experimental results show that GMLDA gets a competitive performance in pose and illumination invariant face recognition and text-image retrieval.

Download English Version:

https://daneshyari.com/en/article/409626

Download Persian Version:

https://daneshyari.com/article/409626

Daneshyari.com