Neurocomputing 171 (2016) 425–433

Contents lists available at ScienceDirect

Neurocomputing

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

Unsupervised discriminant canonical correlation analysis based on spectral clustering

Sheng Wang [a], Jianfeng Lu [a,*], Xingjian Gu [a], Benjamin A. Weyori [a,b], Jing-yu Yang [a]

[a] School of Computer Science and Engineering, Nanjing University of Science and Technology, Nanjing, China
[b] Department of Computer Science and Informatics, University of Energy and Natural Resources, Sunyani, Ghana

ARTICLE INFO

Article history:
Received 6 November 2014
Received in revised form
13 May 2015
Accepted 24 June 2015
Communicated by Xiaoqin Zhang
Available online 6 July 2015

Keywords:
Canonical correlation analysis
Clustering
Feature fusion

ABSTRACT

Canonical correlation analysis (CCA) has been widely applied to information fusion. However, it only considers the correlated information between the paired data and ignores the correlated information between the samples in the same class. Furthermore, class information is helpful for CCA to extract the discriminant feature, but there is no class information available in application of clustering. Thus, it is difficult to utilize the correlated information between the samples in the same class. In order to utilize this correlated information, we propose a method named Unsupervised Discriminant Canonical Correlation Analysis based on Spectral Clustering (UDCCASC). Class membership of the samples is calculated using the normalized spectral clustering, while the mappings for feature fusion are computed by using the generalized eigenvalue method. These two algorithms are executed alternately before the desired result is obtained. Two extensions of UDCCASC are proposed also to deal with multi-view data and nonlinear data. The experimental results on MFD dataset, ORL dataset, MSRC-v1 dataset show that our methods outperform traditional CCA and part of state-of-art methods for feature fusion.

1. Introduction

Up to date, lots of studies on information fusion have been reported in the literature. Generally, they can be divided into three categories [1]: data fusion (low-level fusion), feature fusion (intermediate-level fusion), and decision fusion (high-level fusion). Feature fusion is to obtain new combinations of features, which have more discriminant power for clustering or classification.

Presently, there are many effective feature fusion algorithms [2–7] to extract the discriminant features and reduce the redundant information. For example, Chetty and Wagner [7] proposed a feature fusion method, named feature-level audiovisual fusion, for checking liveness in face–voice person authentication. Canonical Correlation Analysis is one of the most effective and most applied method. It employs two views of the same object to get a pair of vectors (denoted as $\mathbf{W}_x$ and $\mathbf{W}_y$), and makes the mapped data (denoted as $\mathbf{W}_x^T\mathbf{x}$ and $\mathbf{W}_y^T\mathbf{y}$) correlated maximally. Due to its good performance, CCA has been widely used for feature fusion, e.g., the joint use of pixel features and wavelet features for image recognition [8].

Recently, lots of works have been devoted to CCA and many improved variants are proposed. Ridderstolpe et al. [9] utilized CCA to explore the relationship between possible risk factors and several clinical outcomes in cardiac surgery. Hardoon et al. [10] proposed kernel canonical correlation analysis (KCCA) to maintain the nonlinear correlation of the two sets of features. To discover the local manifold structure of the data, Sun et al. [11] proposed Locality Preserved CCA (LPCCA). Furthermore, many supervised variants are proposed. Sharma et al. [12] presented a general framework for feature fusion, and it can be used to extend classical feature techniques into Generalized multi-view ones such as Linear Discriminant Analysis (LDA) [13], Locality Preserving Projection (LPP) [14], Neighborhood Preserving Embedding (NPE) [15] and Marginal Fisher Analysis (MFA) [16]. Rasiwasia et al. [17] proposed mean Canonical Correlation Analysis (mean-CCA) and Cluster Canonical Correlation Analysis (cluster-CCA), which can be applied to the case that the cluster information of each data (or class membership of each data) is available.

Fig. 1 shows three examples of paired data of two categories. These three data pair 1–1′, 2–2′ and 3–3′ can be seen clearly in Fig. 1, while data pair 1–1′ and 2–2′ belong to the same class and 3–3′ comprises the samples of the other class. The correlation between data in the same class can be classified into three categories: (a) the correlation between paired data, which is shown in Fig. 1 by the black dashed lines linking 1-to-1′, 2-to-2′ and 3-to-3′; (b) the correlation between the same class across views and it can be seen from Fig. 1 as a correlation across view

* Corresponding author.
E-mail address: lujf@njust.edu.cn (J. Lu).

http://dx.doi.org/10.1016/j.neucom.2015.06.043
0925-2312/

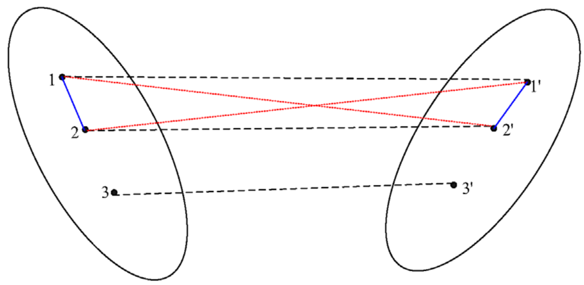

Fig. 1. The correlation between samples in the same class. (For interpretation of the references to color in this figure, the reader is referred to the web version of this paper.)

which is shown by the red dotted lines linking the samples 1-to-2′; (c) the correlation between the same class within views and it can be seen clearly from the figure as the relationship within a class and is represented by the blue solid lines linking 1-to-2 and 1′-to-2′. Given pairs of samples $\{\mathbf{x}_i, \mathbf{y}_i\}_{i=1}^n$, CCA only considers the correlated information of the paired data as the category (a); and it can be seen as unsupervised feature extraction method in nature. The correlated information in category (b) and (c) may contain some important discriminant information, and maximization of these correlations will make mapped samples more compact. Furthermore, it can enhance the performance of classification or clustering. Many supervised variants of CCA [18,12] are proposed and can obtain remarkable performances. However, they only considered the correlation of the category (a) and (b), but ignored that of category (c).

In [19–21], Yang et al. argued that some connection may exist between the classifier and dimension reduction(DR) methods. If the DR method does not match the classifier, the performance of the classification system will be degraded. To connect DR methods with classifiers, one feasible way is to design the DR methods according to the classification rule of a specific classifier. Based on this idea, Yang et al. developed a DR method known as the local mean based nearest neighbor discriminant analysis (LM-NNDA) [19] according to the local mean based on nearest neighbor (LM-NN) classifier [22]. Meanwhile, Chen and Jin [21] also proposed another DR method known as reconstructive discriminant analysis based on linear Regression classification (LRC) [23]. Experimental results showed that maintaining this relation can be helpful to upgrade the performance of these DR methods.

Up to date, CCA is still used in many clustering algorithms, such as [3,4]. However, both of them can be seen as unsupervised methods in nature without considering the correlation of category (b) and (c). To utilize the correlated information between the samples in the same class of categories (a), (b) and (c), we propose a method Discriminant Canonical Correlation Analysis (DCCA) for feature fusion. However, there is no label available before carrying out clustering. Thus, DCCA cannot be directly applied to extract discriminant feature for clustering. As we all know, the cluster membership of each sample can be got after carrying out clustering. Inspired by the DR algorithms according to classifiers, we argue that there may be some connection between the DR methods and the clustering algorithms. Spectral clustering [24] algorithm was proposed by Shi et al., where most recent developed algorithms [25–28] are based on its high performance. In Section 4, we demonstrate by some simple algebraic operations that the calculated results by spectral clustering equals that maximize the correlation between samples of the same class. Therefore, DCCA can be considered as the DR method induced by modified Spectral Clustering. Based on DCCA and SC, we proposed a method named Unsupervised Discriminant Canonical Correlation Analysis based on Spectral Clustering (UDCCASC). Thus, the relation between DR method and clustering algorithms can be persevered in UDCCASC which may be more suitable for clustering. Furthermore, multi-view UDCCASC is developed to cope with multi-view data (more than two views), and kernel UDCCASC is to deal with nonlinear data.

This paper is an extension of our international conference on pattern recognition (ICPR) paper [29]. In this paper, we alter the method for clustering and add the way to balance the weight of correlated information of three categories. Further, we provide the relation between DCCA and normalized spectral clustering. In addition, more experiments are carried out to confirm the effectiveness of UDCCASC.

The rest of this paper is organized as follows. In Section 2, Unsupervised Discriminant Canonical Correlation Analysis based on Spectral Clustering will be described in detail. The multi-view extension and kernel extension will be proposed in Section 3. Section 4 will present in the relationship between discriminant canonical correlation analysis and spectral clustering. The experiments are conducted on three remarkable datasets to test the performance of our method in Section 5. Finally, conclusions are drawn in Section 6.

2. Unsupervised discriminant canonical correlation analysis for two views

This section will present a detailed explanation of the method of discriminant canonical correlation analysis based on spectral clustering. Given pairs of samples $\{\mathbf{x}_{1i}, \mathbf{x}_{2i}\}_{i=1}^n$, $\mathbf{x}_{1i}$ and $\mathbf{x}_{2i}$ are the different descriptions of the same sample and the mean of these two views data are zero. For each symbol of original multiple view data, we define that x_{mn} denotes the n-th data of m view and X_m represents the data matrix of m view. $\mathbf{W_i}$ denotes the mapping for i-th view. Without loss of generality, we use 1-of-K scheme to represent the class membership of samples such as $\mathbf{h_i} \in \Re^{1\times c}$ (c denotes the number of class of training samples). The label of $\{\mathbf{x}_{1i}, \mathbf{x}_{2i}\}_{i=1}^n$ is denoted as one matrix $\mathbf{H} \in \Re^{n\times c}$.

From the above analysis, we know that CCA only considers the correlated information category (a). Correlated information of category (b) and (c) is not considered, which may make CCA very sensitive to noise and apt to over-fitting. Some variants of supervised CCA are proposed [30,31,12], in which authors argued that the correlated information of category (b) should be considered. Further, we argue that the correlation of category (c) should be considered also.

However, in many real applications, there is a little label information available. To utilize the correlated information of category (a), (b) and (c), unsupervised discriminant canonical correlation analysis based on spectral clustering (UDCCASC) is proposed. The object function can be formulated as follows:

$$\{\mathbf{W_1}, \mathbf{W_2}, \lambda, \mathbf{H}\} = \arg\max_{\mathbf{W_1}, \mathbf{W_2}, \lambda, \mathbf{H}} J(\mathbf{W_1}, \mathbf{W_2}, \lambda, \mathbf{H})$$

$$= \arg\max_{\mathbf{W_1}, \mathbf{W_2}, \lambda, \mathbf{H}} \left\{ \beta \sum_i \mathbf{W_1^T} \mathbf{x}_{1i} \mathbf{x}_{2i}^T \mathbf{W_2^T} + (1-\beta) \begin{pmatrix} \lambda_1^r \sum_{i,j} \mathbf{S}_{ij} \mathbf{W_1^T} \mathbf{x}_{1i} \mathbf{x}_{2j}^T \mathbf{W_2^T} \\ +\lambda_2^r \sum_{i,j} \mathbf{S}_{ij} \mathbf{W_1^T} \mathbf{x}_{1i} \mathbf{x}_{1j}^T \mathbf{W_2^T} \\ +\lambda_3^r \sum_{i,j} \mathbf{S}_{ij} \mathbf{W_1^T} \mathbf{x}_{2i} \mathbf{x}_{2j}^T \mathbf{W_2^T} \end{pmatrix} \right\}$$

$$\sum_{i=1}^n \left| \mathbf{W}_1^T \mathbf{x}_{1i} \right|^2 = 1; \quad \sum_{i=1}^n \left| \mathbf{W}_2^T \mathbf{x}_{2i} \right|^2 = 1; \quad \sum_{i=1}^3 \lambda_i = 1 \tag{1}$$

We define class similarity matrix as $\mathbf{S_{ij}} = \mathbf{h_i}\mathbf{h_j^T}$, and $\mathbf{h_i}$ is the i-th row of $\mathbf{H}$. We employ $\lambda = [\lambda_1, \lambda_2, \lambda_3]$ to balance the weight of each category of correlation, which is a popular scheme of controlling the weight of multiple terms[32,33].

The first term in Eq. (1) represents the correlation between the paired data, which is shown in Fig. 1 by the black dashed lines. The second term represents the correlation between the same class across views and it can be seen from Fig. 1 as a correlation across view which