Pattern Recognition 47 (2014) 3907–3919

Contents lists available at ScienceDirect

Pattern Recognition

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

Graph regularized multiset canonical correlations with applications to joint feature extraction

Yun-Hao Yuan [a,*], Quan-Sen Sun [b]

[a] Department of Computer Science and Technology, Jiangnan University, No. 1800, Lihu Street, Wuxi 214122, China
[b] School of Computer Science and Engineering, Nanjing University of Science & Technology, No. 200, Xiaolingwei Street, Nanjing 210094, Jiangsu, China

ARTICLE INFO

Article history:
Received 23 February 2013
Received in revised form
24 March 2014
Accepted 19 June 2014
Available online 30 June 2014

Keywords:
Pattern recognition
Canonical correlation analysis
Multiset canonical correlations
Graph embedding
Feature extraction

ABSTRACT

Multiset canonical correlation analysis (MCCA) is a powerful technique for analyzing linear correlations among multiple representation data. However, it usually fails to discover the intrinsic geometrical and discriminating structure of multiple data spaces in real-world applications. In this paper, we thus propose a novel algorithm, called graph regularized multiset canonical correlations (GrMCCs), which explicitly considers both discriminative and intrinsic geometrical structure in multiple representation data. GrMCC not only maximizes between-set cumulative correlations, but also minimizes local intraclass scatter and simultaneously maximizes local interclass separability by using the nearest neighbor graphs on within-set data. Thus, it can leverage the power of both MCCA and discriminative graph Laplacian regularization. Extensive experimental results on the AR, CMU PIE, Yale-B, AT&T, and ETH-80 datasets show that GrMCC has more discriminating power and can provide encouraging recognition results in contrast with the state-of-the-art algorithms.

1. Introduction

The techniques for joint feature extraction have become popular in recent years for multiple representation data. In many real-world applications in pattern recognition, computer vision, and data visualization, the same objects are usually represented in multiple high-dimensional feature spaces, for example, genes represented by genetic activity feature and text information feature [1], speakers represented by audio feature and visual feature [2]. Since high dimensionality obviously increases the space-time requirements and intractability for processing the data and various feature representations may have very different statistical properties, how to learn meaningful low-dimensional representations from multiple high-dimensional representations is a challenging problem.

Currently, researchers have developed many useful feature extraction or dimensionality reduction techniques for multi-representation data. Among all the methods, canonical correlation analysis (CCA) [3] is undoubtedly the most widely used one for feature extraction and fusion by analyzing linear correlations between two sets of features. In image recognition, Sun et al. [4] employed CCA to first extract low-dimensional correlation features from two representations and then fused them by given strategies for classification tasks. From the viewpoint of regression, Foster et al. [5] performed CCA to derive two sets of low dimensional embeddings and compute the regression function based on these embeddings. However, standard CCA is a linear unsupervised subspace learning method, and thus it is difficult to preserve the discriminative information of data in canonical subspaces. To solve this issue, Sun et al. [6] proposed a discriminant CCA (DCCA) by using within-class and between-class information of training samples. The extracted features are more discriminative for classification tasks. From the point of view of nonlinearity, kernel CCA (KCCA) [7] and locality preserving CCA (LPCCA) [8] were proposed to capture nonlinear correlations between two sets of data. This consideration of the nonlinearity makes KCCA and LPCCA more powerful than CCA for joint feature extraction. Moreover, some other variants of CCA can be found in [9–12].

Recently, a CCA-related method [13], i.e., partial least squares (PLS), was also presented for joint feature extraction from two different representations in face recognition. The PLS-based method can simultaneously project samples into two low-dimensional subspaces, where samples from one representation as regressor and one from the other representation as response. In addition, Sharma et al. [14] presented a general feature extraction framework called generalized multiview analysis (GMA) for two-set high-dimensional data, which can subsume a number of

* Corresponding author. Tel.: +86 25 84315142; fax: +86 25 84318156.
E-mail addresses: yyhzbh@163.com (Y.-H. Yuan), sunquansen@njust.edu.cn (Q.-S. Sun).

http://dx.doi.org/10.1016/j.patcog.2014.06.016

representative dimensionality reduction methods as its special cases, for instance, CCA and PLS. Meanwhile, Sharma et al. [14] also gave an extension of GMA for more than two sets of data. However, the GMA's extension has many parameters,[1] especially when the number of feature representations is very large. This may result in higher computational costs for searching the optimal parameters in numerous real-world applications. In multi-label learning, Yu et al. [15] proposed a multi-output regularized feature projection (MORP) method for dimensionality reduction of multi-label data, in which one view (representation) is derived from the input features and the other view is derived from the output data (i.e., class labels). MORP obtains a low-dimensional feature space that captures both the information of the original feature space and the label space.

However, most of the foregoing methods work well only on two sets of data. When multiple feature representations (at least three ones) arise, they are neither efficient nor optimal for classification tasks [16]. To solve this problem, the idea of multiset canonical correlation analysis (MCCA) [17] has been introduced for joint feature extraction on multiple representation data. MCCA, a generalized extension of CCA, is usually used to reveal linear correlations among more (than two) sets of variables by projecting all sets of variables into respective canonical subspaces. Yuan et al. [18] applied the generalized correlation coefficient to propose a multiset integrated canonical correlation analysis (MICCA) framework. This method projects multiple high-dimensional representations in parallel into respective low-dimensional subspaces, in which multiple sets of features are fused to form effective feature vectors for recognition. As the lacking of discriminant, a discriminative version of MICCA [19] was proposed by using the within-class information of training samples. Hou et al. [20] proposed a novel method called multiple component analysis (MCA) for multi-representation data, which first performs joint feature extraction by a higher-order covariance tensor and then learns orthogonal subspaces through higher-order singular value decomposition (HOSVD) [21] for image recognition tasks.

Different with correlation analysis-based methods, Xia et al. [22] developed a new spectral embedding approach, called multi-view spectral embedding (MSE), which can learn a consensus low-dimensional embedding for multiple feature representations. Since the feature mapping from multiple high-dimensional feature spaces to the low-dimensional subspace is implicit in MSE, it will unavoidably suffer from an out-of-sample problem in pattern classification tasks. That is, it is unclear how to project a new test sample into the low-dimensional subspace. Subsequently, Kan et al. [16] proposed a multiview discriminant analysis (MvDA) approach by simultaneously maximizing between-class variations and minimizing within-class variations of the projected data for robust object recognition. Moreover, some other dimensionality reduction methods [23–27] have also been presented for multi-representation data.

Recently, many studies [28–32] have shown that considering the intrinsic geometrical structure hidden in data can significantly enhance the learning performance of dimensionality reduction methods. In particular, the framework of graph embedding [33] can provide a unified view for a broad set of such algorithms, such as locality preserving projections (LPP) [28,29], Laplacian eigenmap (LE) [30], locally linear embedding (LLE) [31], and isometric feature mapping (ISOMAP) [32]. In other words, these methods can be incorporated into a graph-based learning framework. Motivated by recent progress in correlation analysis and graph embedding, in this paper we propose a novel algorithm for joint feature extraction, called graph regularized multiset canonical correlations (GrMCCs), which explicitly considers both discriminative and intrinsic geometrical structure in multiple representation data. GrMCC not only maximizes between-set cumulative correlations, but also minimizes local intraclass scatter and simultaneously maximizes local interclass separability by using the nearest neighbor graphs on within-set data. Thus, it can leverage the power of both MCCA and discriminative graph Laplacian regularization. Extensive experimental results on five benchmark datasets demonstrate that GrMCC can provide encouraging performance improvements compared with KCCA, MCA, MCCA, MICCA, and baseline algorithms PCA and CCA.

[1] If there are m feature representations, then the number of parameters is $(m-1)(m+4)/2$.

2. Background and related work

The work most related to our proposed method is CCA, MCCA, and graph embedding. Therefore, in this section, we provide a brief description for them.

2.1. Canonical correlation analysis

In CCA, given two zero-mean random vectors $x \in R^p$ and $y \in R^q$, the objective of CCA is to compute a pair of projection directions, $\alpha \in R^p$ and $\beta \in R^q$, such that the correlation of canonical variates $\alpha^T x$ and $\beta^T y$ is maximized by

$$\rho(\alpha, \beta) = \frac{E(\alpha^T x y^T \beta)}{\sqrt{E(\alpha^T x x^T \alpha) \cdot E(\beta^T y y^T \beta)}} = \frac{\alpha^T S_{xy} \beta}{\sqrt{\alpha^T S_{xx} \alpha \cdot \beta^T S_{yy} \beta}}, \quad (1)$$

where $E(\cdot)$ denotes the expectation, S_{xx} and S_{yy} are respectively, within-set covariance matrices of vectors x and y, and S_{xy} is a between-set covariance matrix between vectors x and y. Clearly, the canonical correlation criterion in (1) is affine-invariant to the arbitrary scaling of α and β. According to this characteristic, CCA needs to normalize the canonical transformations α and β by setting

$$\alpha^T S_{xx} \alpha = 1 \quad \beta^T S_{yy} \beta = 1. \quad (2)$$

On the foregoing basis, the first pair of projection directions, α_1 and β_1 are computed by maximizing the criterion (1) with constraints (2). After this, the kth pair of projection directions, α_k and β_k, where $2 \le k \le r$ and $r = rank(S_{xy})$, are found by continually maximizing the criterion (1) with the following constraints:

$$\begin{cases} \alpha^T S_{xx} \alpha = 1, & \beta^T S_{yy} \beta = 1, \\ \alpha_j^T S_{xx} \alpha = 0, & \beta_j^T S_{yy} \beta = 0, \quad (j = 1, 2, \cdots, k-1). \end{cases} \quad (3)$$

2.2. Multiset canonical correlation analysis

MCCA is an important technique which can analyze linear relationships between multiple sets of random variables. At present, MCCA has many different forms [17]. Thereinto, the following form (i.e., standard MCCA) is a natural and direct extension of CCA.

Given m sets of zero-mean random vectors $\{x_i \in R^{d_i}\}_{i=1}^m$, a set of projection directions $\{\alpha_i \in R^{d_i}\}_{i=1}^m$, called multiset canonical transformations (MCTs), is found to maximize the sum of pair-wise correlations between multiset canonical variates (MCVs) $\{\alpha_i^T x_i\}_{i=1}^m$ as

$$\rho(\tilde{\alpha}) = \sum_{i=1}^m \sum_{j=1}^m \alpha_i^T S_{ij} \alpha_j, \quad (4)$$

where $\tilde{\alpha}^T = (\alpha_1^T, \alpha_2^T, \ldots, \alpha_m^T)$, S_{ii} is the within-set covariance matrix of vector x_i, and $S_{ij} (i \ne j)$ is the between-set covariance matrix between vectors x_i and x_j.