Neurocomputing ∎ (∎∎∎∎) ∎∎∎–∎∎∎

Contents lists available at ScienceDirect

Neurocomputing

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

A unified multiset canonical correlation analysis framework based on graph embedding for multiple feature extraction

XiaoBo Shen [a], QuanSen Sun [a,*], YunHao Yuan [b]

[a] School of Computer Science and Engineering, Nanjing University of Science & Technology, Nanjing 210094, China
[b] School of Internet of Things, Jiangnan University, Wuxi 214122, China

ARTICLE INFO

Article history:
Received 22 January 2014
Received in revised form
3 April 2014
Accepted 2 June 2014
Communicated by Haowei Liu

Keywords:
Multiset canonical correlation analysis
Graph embedding
Multiple feature extraction
Feature fusion
Dimensionality reduction
Discriminant analysis

ABSTRACT

Multiset canonical correlation analysis (MCCA) can simultaneously reduce the dimensionality of multimodal data. Thus, MCCA is very much suitable and powerful for multiple feature extraction. However, most existing MCCA-related methods are unsupervised algorithms, which are not very effective for pattern classification tasks. In order to improve discriminative power for handling multimodal data, we, in this paper, propose a unified multiset canonical correlation analysis framework based on graph embedding for dimensionality reduction (GbMCC-DR). Under GbMCC-DR framework, three novel supervised multiple feature extraction methods, i.e., GbMCC-LDA, GbMCC-LDE, and GbMCC-MFA are presented by incorporating several well-known graphs. These three methods consider not only geometric structure of multimodal data but also separability of different classes. Moreover, theoretical analysis further shows that, in some specific circumstances, several existing MCCA-related algorithms can be unified into GbMCC-DR framework. Therefore, this proposed framework has good expansibility and generalization. The experimental results on both synthetic data and several popular real-world datasets demonstrate that three proposed algorithms achieve better recognition performance than existing related algorithms, which is also the evidence for effectiveness of GbMCC-DR framework.

1. Introduction

Feature extraction is an important research topic in computer vision and pattern recognition. Most feature extraction methods including principal component analysis (PCA) [1] and linear discriminant analysis (LDA) [2] belong to single modal techniques and thus can only handle such data from single view or single channel. However, multimodal data, e.g., sound and images from the same video, text and pictures from the same webpage is prevalent in practice. Therefore, how to deal with such multimodal data simultaneously is a fundamental and practical problem. One popular approach to address this problem is feature fusion technology [3]. The advantage of feature fusion is obvious: firstly, it keeps the effective discriminant information of multiple features; secondly, it also eliminates the redundant information to certain degree. Consequently, feature fusion is significant for recognition and has received more and more attentions in pattern recognition [4–6].

Classical feature fusion approaches simply concatenate or integrate multiple features together by using serial feature fusion [4] or parallel feature fusion strategies [5,6]. Although these methods improve classification performance, the main drawback is that they hardly reflect the intrinsic correlation between different features, which is helpful to fusion and recognition. In order to solve this problem, Sun et al. [7] firstly applied canonical correlation analysis (CCA) [8,9] into feature fusion field. CCA maximizes the correlations between two sets of features obtained from the same patterns, and then extracts the effective discriminant feature vectors. Thus CCA obviously achieves better performances than previous fusion methods. In essence, CCA is an unsupervised linear feature extraction method. Recently several variants of CCA have been developed [7,10–15]. In order to improve discriminative power of CCA features, several supervised algorithms [10,11,13], e.g., generalized CCA (GCCA) [10], discriminant CCA (DCCA) [11] were proposed. They incorporate label information into CCA from different angles, and improve multimode recognition rates to some extent. However, at present, most CCA-related methods only focus on two sets of features, which cannot properly depict the correlations among several (more than two) sets of features.

Multiset canonical correlation analysis (MCCA) [16,17] is a powerful method to analyze linear relationships among several (more than two) sets of features, and thus is a generalized extension of CCA in essence. MCCA obtains high correlations between several feature sets simultaneously by optimizing the

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

http://dx.doi.org/10.1016/j.neucom.2014.06.015
0925-2312/

characteristics of dispersion matrix of the transformed features. According to different criteria and constraints, several models about MCCA have been developed [16–19]. Recently, based on generalized correlation coefficient, Yuan et al. [18] proposed a novel multiset integrated canonical correlation analysis framework (MICCA). MICCA fuses several feature sets and shows its effectiveness in handwritten numerals classification. Meanwhile, Jing et al. [19] presented a color image CCA method (CICCA) for color image recognition. CICCA extracts canonical correlated features from three color components of color images and provides an analytical solution by solving three eigenproblems simultaneously. Experimental results on several color face image databases demonstrate its better performance than several single modal related methods.

Graph embedding [20,21] is a popular and effective framework for dimensionality reduction (DR). It focuses on modeling the pairwise relationships among data and utilizes graph-based structures. Graph embedding provides a general formulation to unify many existing DR algorithms, include supervised ones [2,20,22,23], semi-supervised ones [24,25], and unsupervised ones [1,26,27]. On the other hand, graph embedding can also be utilized as a tool to develop some new DR algorithms, e.g., marginal fisher analysis (MFA) [20] and matrix-based marginal fisher analysis (MMFA) [28]. Inspired by some graph-based methods [26], some variants of CCA have been proposed. Locality preserving CCA (LPCCA) [12] incorporates local geometry into objective of CCA. By constructing the locally linear correlation defined in a local field, LPCCA discovers the local manifold structure of data, and achieves good performance in pose estimation and data visualization. In essence, LPCCA is an unsupervised feature extraction method. Later, local discrimination CCA (LDCCA) [13] is proposed to consider the combination of local properties and discrimination between different classes. LDCCA maximizes local within-class correlations and minimizes local between-class correlations simultaneously to achieve good recognition performance. From the above, we find that these methods mainly focus on handling two-view data. Then how to present a unified formulation to embed prior information such as geometric structure, label information into multiple (more than two) feature extraction still remains unknown.

Inspired with the success of graph embedding, we propose a unified multiset canonical correlation analysis framework based on graph embedding (GbMCC-DR). With an equivalent formulation of MCCA, we naturally embed some prior information of multimodal data into the objective of MCCA. On the other hand, most existing MCCA-related methods, e.g., MCCA, CICCA, and MICCA are unsupervised algorithms, which are not very effective for pattern classification tasks. In order to improve discriminative power for handling multimodal data, based on three popular single-modal DR algorithms, i.e., LDA [2], local discriminant embedding (LDE) [22], and MFA [20], we develop three novel supervised multiple feature extraction methods, termed GbMCC-LDA, GbMCC-LDE, and GbMCC-MFA, respectively. These three proposed algorithms introduce label information as well as geometric structure into GbMCC-DR framework by incorporating different types of graphs, and all extract more discriminative features than existing MCCA-related methods.

The rest of this paper is organized as follows. In Section 2, we briefly describe basic theories of MCCA and graph embedding. In Section 3, GbMCC-DR framework and three novel supervised multiple feature extraction methods are proposed. Moreover, we further discuss the relationships between GbMCC-DR framework and some existing related works. In Section 4, we experimentally evaluate these three proposed supervised algorithms both on synthetic data and several real-world datasets. Finally, concluding remarks and some discussions of future work will be given in Section 5.

2. Some related works

In this section, we give a brief review of basic theories of MCCA as well as graph embedding.

2.1. Multiset canonical correlation analysis

Multiset canonical correlation analysis (MCCA) [16,17] is a technique for analyzing linear relationships between several (more than two) sets of features. In essence, it is a natural extension of CCA. In mathematics, according to different objectives and constraints, there are several different models [17] of MCCA. Among them, the following model, termed as SUMCOR [17], has been widely used for its simplicity and analytical solution.

Suppose that m random vectors are $x^{(i)} \in \Re^{p_i}$ $(i=1,2,\ldots,m)$, and all the variables are centered, i.e., $E(x^{(i)})=0$. MCCA seeks m projection directions $\alpha^{(i)} \in \Re^{p_i}$, to maximize the sum of the pairwise correlation between transformed variables $\alpha^{(i)T}x^{(i)}$ under some constraint. Specifically, the optimization problem of MCCA is as follows:

$$\max \rho(\alpha^{(1)},\alpha^{(2)},\ldots,\alpha^{(m)}) = \sum_{i=1}^{m}\sum_{j\neq i,j=1}^{m} \alpha^{(i)T}S_{ij}\alpha^{(j)}$$

$$\text{s.t.}\quad \sum_{i=1}^{m}\alpha^{(i)T}S_{ii}\alpha^{(i)} = 1 (i=1,2,\ldots,m) \tag{1}$$

where S_{ii} is within-set covariance matrix of random variable $x^{(i)}$, S_{ij} $(i\neq j)$ is between-set covariance matrix between random variable $x^{(i)}$ and $x^{(j)}$.

With Lagrangian multiplier method, the solution of MCCA can be equally transformed into a generalized eigenvalue problem (refer [17] for some details). It is worth mentioning that when MCCA only handles two sets of variables (i.e., $m=2$), it will obviously reduce to CCA.

2.2. Graph embedding

Graph embedding [20,21], as a general DR framework, provides a new perspective to understand many existing DR algorithms. It focuses on modeling the pairwise relationships among data and utilizes graph-based structures. In graph embedding, the key problem is to construct proper graphs to characterize desired statistical or geometric properties of data sets.

Assume training sample set is given as $X=[x_1,x_2,\ldots,x_N] \in \Re^{p\times N}$, where N is the number of training samples, and p is the feature dimensionality. In graph embedding, intrinsic graph $G=\{X,W\}$ is an undirected weighted graph, where $W=[W_{ij}] \in \Re^{N\times N}$ is similar matrix. It is used to record the edge weights that characterize the similarity relationships between samples. The objective of graph embedding is to seek optimal projection direction w, which satisfies

$$\min_{w} \sum_{i\neq j} \| w^T x_i - w^T x_j \|^2 W_{ij} = w^T X L X^T w$$

$$\text{s.t.}\quad w^T X B X^T w = 1 \text{ or } w^T X L^P X^T w = 1 \tag{2}$$

where $L=D-W$ is graph Laplacian of G, $D=\mathrm{diag}(D_{11},D_{22},\ldots,D_{NN})$, and D_{ij} is the ith row (or column) sum of similar matrix W. Depending on the property of a problem, one of the two constraints in (2) will be used in the optimization. The first constraint is chosen for scale normalization, where B is a diagonal matrix. Otherwise, penalty graph $G^P=\{X,W^P\}$ is required for second constraint, where L^P, W^P are graph Laplacian and similarity matrix of G^P, respectively. Yan et al. [20] show that, by specifying W and B (or W^P), many existing DR algorithms [1,2,22,23,25,26,29] can be unified into graph embedding.

Please cite this article as: X. Shen, et al., A unified multiset canonical correlation analysis framework based on graph embedding for multiple feature extraction, Neurocomputing (2014), http://dx.doi.org/10.1016/j.neucom.2014.06.015