Information Fusion 41 (2018) 119–128

Contents lists available at ScienceDirect

Information Fusion

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

Multiview dimension reduction via Hessian multiset canonical correlations

Weifeng Liu[a,*], Xinghao Yang[a], Dapeng Tao[b,*], Jun Cheng[c,d], Yuanyan Tang[e,f]

[a] College of Information and Control Engineering, China University of Petroleum (East China), Qingdao 266580, Shandong, China
[b] School of Information Science and Engineering, Yunnan University, Kunming 650091, Yunnan, China
[c] Shenzhen Key Lab for CVPR, Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen, Guangdong, China
[d] The Chinese University of Hong Kong, Hong Kong, China
[e] Faculty of Science and Technology, University of Macau, Macau 999078, China
[f] College of Computer Science, Chongqing University, Chongqing 400000, China

ARTICLE INFO

Article history:
Received 20 January 2017
Revised 10 August 2017
Accepted 3 September 2017
Available online 5 September 2017

Keywords:
Multiview
Dimension reduction
Hessian
Canonical correlation analysis

ABSTRACT

Canonical correlation analysis (CCA) is a main technique of linear subspace approach for two-view dimension reduction by finding basis vectors with maximum correlation between the pair of variables. The shortcoming of the traditional CCA lies that it only handles data represented by two-view features and cannot reveal the nonlinear correlation relationship. In recent years, many variant algorithms have been developed to extend the capability of CCA such as discriminative CCA, sparse CCA, kernel CCA, locality preserving CCA and multiset canonical correlation analysis (MCCA). One representative work is Laplacian multiset canonical correlations (LapMCC) that employs graph Laplacian to exploit the nonlinear correlation information for multiview high-dimensional data. However, it possibly leads to poor extrapolating power because Laplacian regularization biases the solution towards a constant function. In this paper, we present Hessian multiset canonical correlations (HesMCC) for multiview dimension reduction. Hessian can properly exploit the intrinsic local geometry of the data manifold in contrast to Laplacian. HesMCC takes the advantage of Hessian and provides superior extrapolating capability and finally leverage the performance. Extensive experiments on several popular datasets for handwritten digits classification, face classification and object classification validate the effectiveness of the proposed HesMCC algorithm by comparing it with baseline algorithms including TCCA, KMUDA, MCCA and LapMCC.

1. Introduction

Many computer vision tasks employ multiple information rather than a single representation to achieve more robust performance. For example, an image can be described with color, shape, and texture features [1–3]. The extracted visual features usually have high dimensions of up to hundreds or thousands, which often causes the problem called the curse-of-dimensionality [4,5]. Hence multiview dimension reduction algorithms [6,7] have been subsequently proposed with the purpose of finding an appropriate low-dimensional feature subspace from multiview high-dimensional features. Canonical correlation analysis (CCA) [8] is one of the most representative techniques and has been widely applied to many multiview learning applications including classification, retrieval, regression and clustering.

Canonical correlation analysis (CCA) proposed by Hotelling [8] seeks a pair of linear transformation for two view high-dimensional features such that the corresponding low-dimensional projections are maximally correlated. In recent years, dozens of CCA extensions have been developed and these algorithms can be roughly categorised into the following four groups namely discriminative CCA, sparse CCA, kernel CCA and locality preserving CCA.

Discriminative CCA [9–11] considers the combination of within-class/between-class information and correlated information of training samples to improve the discrimination of the low-dimensional subspace. For example, Kim et al. [9] developed a linear discriminant function of CCA similarly to linear discriminant analysis (LDA) by maximizing the canonical correlations of within-class sets and minimizing the canonical correlations of between-class sets. Wang et al. [10] proposed an unsupervised discriminant CCA based on spectral clustering by utilizing the correlation information between the samples in the same class including the correlation between paired data, the correlation across views, and the correlation within views. Sun et al. [11] added the uncorrelation

* Corresponding author.
E-mail addresses: liuwf@upc.edu.cn (W. Liu), dptao@ynu.edu.cn (D. Tao).

http://dx.doi.org/10.1016/j.inffus.2017.09.001
1566-2535/

constraint into multiview linear discriminant analysis (MLDA) and proposed the multiview ULDA (MULDA) to seek the project directions with minimum redundancy. Besides, they also extended the MULDA to its kernel version, i.e., kernel multiview uncorrelated discriminant analysis (KMUDA), to tackle the nonlinear case.

Sparse CCA [12–14] seeks two sparse canonical vectors to maximize the correlation between the two views. Hardoon and Shawe-Taylor [12] proposed a sparse CCA by minimizing the number of features while maximizing the correlation between the primal view and the dual view. Witten et al. [13] employed penalized matrix decomposition to optimize l_1-regularized sparse CCA. And Chen et al. [14] developed a structured sparse CCA by incorporating the structured-sparsity-inducing penalty to exploit both pre-given group structure and unknown group structure.

Kernel CCA [15–18] finds maximally nonlinear correlations via an implicit nonlinear mapping to increase the flexibility of the feature selection. Lai and Fyfe [15] used a kernel trick to map the data to a high-dimensional feature space and then performed the traditional CCA on the feature space. Akaho [17] developed regularized kernel CCA to avoid the overfitting by incorporating regularization technique into kernel CCA. Zhu et al. [18] utilized a mixture of kernels rather than a single kernel as the nonlinear mapping and improved both interpolation ability and extrapolation ability of kernel CCA.

Locality preserving CCA incorporates locality information into CCA and then preserves the local manifold structure [19] while obtaining the canonical correlation [20]. Sun and Chen [21] decomposed the global nonlinear structure into many locally linear ones, then conducted linear CCA on each small neighborhood and finally integrated the local sub-problems to obtain the canonical correlation.

Although CCA and its variants have achieved practical success, they are in nature a two-view dimension reduction techniques and cannot handle the problem of multiview dimension reduction that often exists in many real-world applications. Multiset canonical correlation analysis (MCCA) [22] then has been developed to tackle the correlation analysis of multiple variables. Luo et al. [23] proposed tensor CCA to maximize the correlation between the multiple canonical variables by finding the best rank-1 approximation of the high-order covariance tensor. And recently, Yuan et al. [24] proposed Laplacian multiset canonical correlations (LapMCC) to discover the nonlinear correlations among multiview features by combining many locally within-view and between-view correlations together.

In spite of LapMCC has achieved promising performance compared with the traditional MCCA algorithms, it has been identified that Laplacian will bias the solution towards a constant function and then lead to poor extrapolating power [25]. On the other hand, Hessian predicts a high order derivatives and has a richer null space, which makes it can steer the prediction varying smoothly along the underlying manifold. In this paper, we integrate Hessian into the multiset canonical correlations and derive Hessian multiset canonical correlations (HesMCC). HesMCC takes the advantage of Hessian and provides superior extrapolating capability. Therefore, HesMCC can significantly leverage the performance. Fig. 1 graphically demonstrates the whole procedure of the proposed HesMCC. Briefly speaking, the contribution of this paper includes the following three folds: (1) we derive the locality preserving canonical correlation by using Hessian; (2) we further formulate HesMCC to analyze multiset canonical correlation; (3) we present the algorithm of HesMCC and conduct extensive experiments to verify the proposed HesMCC.

Finally, we carefully implement HesMCC for multiview dimension reduction and conduct extensive experiments on USPS database for handwritten digits classification, Yale-B database and ChokePoint database for face recognition and ETH-80 database for object classification respectively. We also compare HesMCC with TCCA, KMUDA, MCCA and LapMCC algorithms to evaluate the performance of HesMCC. The experimental results verify the effectiveness of HesMCC by comparison with the baseline algorithms.

The rest of this paper is organized as follows. We firstly review some related works in Section 2. Then we derive the proposed HesMCC algorithm in Section 3. Section 4 details the implementation of HesMCC. And experimental results are discussed in Section 5, followed by the conclusion in Section 6.

2. Related work

In this section, we first briefly summarize the related work for the multiview canonical correlation analysis including multiset canonical correlation analysis (MCCA), multiset integrated canonical correlation analysis (MICCA), tensor canonical correlation analysis (TCCA), and Laplacian multiset canonical correlation (LapMCC). And then, we give a brief null space description of Laplacian and Hessian.

2.1. Traditional multiview CCA related works

Suppose we are given a dataset of n examples with m view representations i.e. $\mathcal{S} = \{x_k^{(1)}, x_k^{(2)}, \cdots, x_k^{(m)}\}_{k=1}^n$, where $x_k^{(i)} \in R^{d_i}$ is the i^{th} view representation of the k^{th} example, d_i is the dimension of the i^{th} view feature as mentioned above.

MCCA [22] seeks a set of linear projection directions $\{\alpha_i \in R^{d_i}\}_{i=1}^m$ such that the sum of pairwise correlations $\{\alpha_i^T x_i\}_{i=1}^m$ is largest, which is formulated as

$$\max\ \rho(\tilde{\alpha}) = \sum_{i=1}^{m}\sum_{j=1}^{m}\alpha_i^T S_{ij}\alpha_j$$
$$\text{s.t.}\ \alpha_i^T S_{ii}\alpha_i = 1,\ \ i = 1, 2, \cdots, m.$$

where $S_{ij} = x_i x_j^T$.

MICCA [26] maximizes the generalized relation coefficient between the multiset example by defining a generalized uncorrelation coefficient $\triangle(\cdot)$

$$\max\ \rho(\tilde{\alpha}) = \sqrt{1 - \triangle^2\left(\alpha_1^T X^{(1)},\ \alpha_2^T X^{(2)}, \cdots, \alpha_m^T X^{(m)}\right)}$$

where $\triangle\,(x_1, x_2, \cdots, x_p) = \frac{\sqrt{det(G(x_1,x_2,\cdots,x_p))}}{\|x_1\|\cdot\|x_2\|\cdots\|x_p\|}$ and $G(\cdot)$ denotes a Gram matrix.

TCCA [23] directly maximizes the canonical correlation of multiviews by straightforwardly analyzing the covariance tensor of the different views

$$\max\ \rho(h_p) = C_{12\cdots m}\bar{\times}_1 h_1^T \bar{\times}_2 h_2^T \cdots \bar{\times}_m h_m^T$$
$$\text{s.t.}\ h_p^T C_{pp} h_p = 1,\ \ p = 1, 2, \cdots, m.$$

where $C_{12\ldots m}$ is the covariance tensor among all views and h_p is the canonical vector.

LapMCC [24] considers local within-view and local between-view correlations by using nearest neighbor graphs

$$\max\ \rho(\tilde{\alpha}) = \sum_{i=1}^{m}\sum_{j=1}^{m}\alpha_i^T S_{ij}^L\alpha_j$$
$$\text{s.t.}\ \alpha_i^T S_{ii}^L\alpha_i = 1,\ \ i = 1, 2, \cdots, m.$$

here $S_{ij}^L = \frac{1}{n^2}X^{(i)}L^{(ij)}X^{(j)T}$ is the within-view $(i = j)$ or between-view $(i \neq j)$ covariance matrix, and $L^{(ij)}$ is the graph Laplacian to characterise the local geometry of example distribution.