Neurocomputing 154 (2015) 257–275

Contents lists available at ScienceDirect

Neurocomputing

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

Globalized and localized canonical correlation analysis with multiple empirical kernel mapping

Changming Zhu, Zhe Wang*, Daqi Gao*

Department of Computer Science & Engineering, East China University of Science & Technology, Shanghai 200237, PR China

ARTICLE INFO

Article history:
Received 5 August 2014
Received in revised form
7 October 2014
Accepted 26 November 2014
Communicated by Xiaojie Su
Available online 8 December 2014

Keywords:
Multiple kernel learning
Empirical kernel mapping
Canonical correlation analysis
Global and local structure

ABSTRACT

Canonical Correlation Analysis (CCA) reveals linear correlation relationship between two feature sets, but fails to discover nonlinear relationship. Kernel CCA (KCCA) overcomes such a shortcoming. Unfortunately, both of them fail to discover local structure of features whereas Locality Preserving CCA (LPCCA) possesses this ability. It is found that LPCCA ignores relationship between global and local structures of features. Moreover, these CCA-based methods have no ability to deal with single-view data which only has single feature set. To this end, we apply Multiple Explicitly Kernel Mapping (MEKM) to the application at first and take global and local structures of features into account. The proposed method is named Globalized and Localized CCA with MEKM (GLCCA-MEKM). Experiments validate that (i) introducing MEKM can map original features into multiple feature spaces so that multiple feature sets of data are obtained. Further in these feature spaces, nonlinear correlation relationship between features are also gotten; (ii) taking global and local structures of features into account makes the mapped features keep both original global and local properties. These processes make GLCCA-MEKM possess the ability to deal with single-view data and be locality-preserving. Therefore, GLCCA-MEKM has below contributions. First, GLCCA-MEKM can inherit the advantages of traditional MEKM, deal with single-view data, and reveal nonlinear correlation relationship between two feature sets. Second, GLCCA-MEKM extracts both global and local structural information more reasonably and coordinates their relationship well. In doing so, mapped features can keep original global and local properties. Finally, classifiers with GLCCA-MEKM obtain better classification performances.

1. Introduction

Suppose a data set $Z=\{z_1, z_2, \ldots, z_n\}$ has n samples and each sample z_i consists of one feature pair, namely $z_i=\{x_i, y_i\}$ where x_i and y_i are two features of this feature pair. Then we say $Z=\{X, Y\}$ and $X=\{x_1, \ldots, x_n\}$ and $Y=\{y_1, \ldots, y_n\}$ are two feature sets. Here, each set is treated as one view of data. When X and Y are two mean-normalized feature sets, as a major linear subspace approach to dimensionality reduction, Canonical Correlation Analysis (CCA) [1,2] aims to find basis vector pair (ω_x, ω_y) in order to make the correlation between the canonical component pair $(\omega_x^T X, \omega_y^T Y)$ maximized. Component ω_x^{TX} (ω_y^{TY}) can be treated as a new form of X (Y) after dimensionality reduction. In other words, we can reduce dimensionality of x_i (y_i) by the projection $\omega_x^T x_i$ $(\omega_y^T y_i)$ and $\omega_x^T x_i$ $(\omega_y^T y_i)$ is a new form of x_i (y_i) in a low-dimensional space. Moreover, fusing ω_x^{TX} and ω_y^{TY} can form a new fused data set which consists of new fused samples $z_{inew}=\{\omega_x^T x_i, \omega_y^T y_i\}$ where $i=1,2,\ldots,n$. Since each data set has not only linear correlation relationship but also nonlinear one between two feature sets, it is found that CCA only reveals linear one whereas ignores nonlinear one. In order to overcome such a problem, Kernel CCA (KCCA) [3] has been proposed. In KCCA, the original features are mapped into higher (even infinite) dimensional spaces which are also named feature spaces via implicit nonlinear mappings. Then a nonlinear problem in the original space is transformed into another more possibly linear one in the feature spaces so as to discover the nonlinear correlation hidden between original features [4]. Experimental results have also validated that compared with CCA, a classifier with KCCA has a better performance than CCA used [3,5]. While both CCA and KCCA extract global structures of features and fail to discover local structures. Thus Locality Preserving CCA (LPCCA) [4] has been proposed to extract local structures. LPCCA takes local neighborhood structures of features into account and captures canonical correlation between feature pairs. Different from CCA and KCCA, LPCCA does not need two mean-normalized feature sets for experiments. Moreover, the experiment has validated that LPCCA performs better than both CCA and KCCA in pose estimation [4]. Although LPCCA overcomes disadvantages of CCA and KCCA, it ignores the relationship between global and local structures of features. Moreover, for CCA, KCCA, and LPCCA, they resort to multiple feature sets and have no ability to deal with single-view data which only has single feature set.

* Corresponding authors.
E-mail addresses: aaronmike@163.com, wangzhe@ecust.edu.cn (Z. Wang), gaodaqihaha@163.com, gaodaqi@ecust.edu.cn (D. Gao).

http://dx.doi.org/10.1016/j.neucom.2014.11.066
0925-2312/

According to above disadvantages of these CCA-based methods, we adopt Multiple Kernel Learning (MKL) which can map original features into multiple feature spaces. As a traditional kernel-based methods, MKL [6–12] maps original feature x into different feature spaces $\mathcal{F}_l, l=1,\ldots,M$, namely $\Phi_l(x): x \rightarrow \mathcal{F}_l$ [9]. Mapped features in a feature space can be treated as a view of data or a feature set. In practice, mapping $\Phi(x)$ has two kinds of forms. One is named Implicit Kernel Mapping (IKM) which is denoted as $\Phi^i(x)$ and the other is named Empirical Kernel Mapping (EKM) which is denoted as $\Phi^e(x)$. MKL with Multiple Implicit Kernel Mapping (MIKM) and Multiple Empirical Kernel Mapping (MEKM) is named Multiple Implicit Kernel Learning (MIKL) [9–18] and Multiple Empirical Kernel Learning (MEKL) respectively. As we know, $\Phi^i(x)$ is supposed to avoid the curse of dimensionality and $\Phi^e(x)$ gives the forms of the mapped features in the feature space directly and this makes the classifier design more easier. Both of them make MKL generate multiple views of data and keep the linearity property of learning machine in the feature space at the same time. For the convenience of dealing with the mapped features directly, we focus on MEKM and take both global and local structural information of features into account so that a new CCA-based method is proposed. We name this new method as Globalized and Localized Canonical Correlation Analysis with Multiple Empirical Kernel Mapping (GLCCA-MEKM). The process of GLCCA-MEKM includes two steps. The first one is using MEKM to map original features into multiple feature spaces in order to deal with single-view or multiple-view data and reveal nonlinear correlation relationship between feature pairs. The second one is getting the projections or new forms of the original features so as to get fused samples with global and local structures of features taken into account.

In fact, GLCCA-MEKM is different from the existing CCA-based methods. First, CCA only reveals linear correlation relationship between two feature sets and fails to discover nonlinear one. But GLCCA-MEKM uses MEKM to form multiple feature spaces and has an ability to deal with a nonlinear problem in the original space. Then GLCCA-MEKM can discover the nonlinear correlation hidden between the original features. Second, KCCA has an ability to reveal both linear and nonlinear correlation relationships but ignores local structures of features. GLCCA-MEKM overcomes disadvantages of KCCA and extracts local properties of features. Third, LPCCA can discover local structures of features but ignores the relationship between global and local structures. To this end, GLCCA-MEKM takes both global and local structures of features into account so that global and local properties of features can be kept. Moreover, CCA, KCCA, and LPCCA resort to multiple feature sets and have no ability to deal with single-view data while GLCCA-MEKM overcomes it. In generally speaking, the novelty of the proposed GLCCA-MEKM lies on the fact that (1) the original features are mapped into multiple feature spaces explicitly by MEKM. So GLCCA-MEKM inherits the advantages of the traditional MEKM and reveals nonlinear correlation relationship between multiple feature sets; (2) GLCCA-MEKM extracts both the global and local structural information and coordinates the relationship between them well. Furthermore, mapped features keep their original global and local properties; (3) classifiers with GLCCA-MEKM have better classification performances.

The rest of this paper is organized as below. In Section 2, for the reason that we use CCA as the base of GLCCA-MEKM, some CCA-based methods are reviewed. Section 3 gives the architecture of the proposed GLCCA-MEKM. Section 4 shows all experimental results. Finally, the conclusions and future work are given in Section 5.

2. Related work

As we said before, CCA [1,2] is the base of the proposed GLCCA-MEKM and reveals linear correlation relationship between two feature sets. But CCA cannot reveal nonlinear correlation relationship whereas KCCA can. Both CCA and KCCA have no ability to discover local structures of features. So LPCCA is proposed to overcome this shortcoming. In this section, we give a brief description of CCA, KCCA, and LPCCA.

2.1. CCA

As one of the principal subspace approaches to dimensionality reduction, CCA aims to find basis vector pair, (ω_x, ω_y), for two mean-normalized feature sets $X=\{x_1,\ldots,x_n\}$ and $Y=\{y_1,\ldots,y_n\}$ such that correlation between canonical component pair $(\omega_x^T X, \omega_y^T Y)$ is maximized [1,2]. Here, in a feature pair $\{x_i, y_i\}$, $x_i \in \mathbb{R}^p$ and $y_i \in \mathbb{R}^q$, n is the number of feature pairs. For CCA, it should solve the below optimization problem with constraints to get ω_x and ω_y:

$$\max \quad \omega_x^T X P_c Y^T \omega_y \qquad \text{s.t.} \begin{cases} \omega_x^T X P_c X^T \omega_x = 1 \\ \omega_y^T Y P_c Y^T \omega_y = 1 \end{cases} \tag{1}$$

where $P_c = I-(1/n)I_n I_n^T$, $I_n=[1,\ldots,1]^T \in \mathbb{R}^n$. I is an $n\times n$ identity matrix. P_c is a mean-normalization matrix which satisfies $P_c^T = P_c$, $P_c^T P_c = P_c$.

By solving this optimization problem (1), we can obtain the below generalized eigen-problem:

$$\begin{pmatrix} & XP_cY^T \\ YP_cX^T & \end{pmatrix}\begin{pmatrix}\omega_x\\ \omega_y\end{pmatrix} = \lambda \begin{pmatrix} XP_cX^T & \\ & YP_cY^T \end{pmatrix}\begin{pmatrix}\omega_x\\ \omega_y\end{pmatrix} \tag{2}$$

where the eigenvalue λ is the objective value to be optimized in (1), namely the canonical correlation. By solving Eq. (2), we can get d eigenvalues which are larger than zero. So d pairs of (ω_x, ω_y) are gotten where the size of ω_x is $p\times 1$ and the size of ω_y is $q \times 1$. Let $\mathcal{W}_x = [\omega_{x1},\ldots,\omega_{xd}]$ and $\mathcal{W}_y = [\omega_{y1},\ldots,\omega_{yd}]$ where ω_{xi} and ω_{yi} form i-th basis vector pair $(\omega_{xi}, \omega_{yi})$. Then $\mathcal{W}_x$ and $\mathcal{W}_y$ can be used as two projective matrices whose columns correspond to the first d largest common eigenvalues of (2). The dimensionality reduction of the original features can be performed in the form of $\mathcal{W}_x^T X^c$ and $\mathcal{W}_y^T Y^c$. Here, $X^c = XP_c$ and $Y^c = YP_c$ are two mean-normalized feature matrices. By this process, the original features can be both reduced to d-dimensional ones. Then it can also be validated that we can use two linear combination of X^c and Y^c to denote the ω_x and ω_y as below:

$$\begin{aligned} \omega_x &= X^c\alpha \\ \omega_y &= Y^c\beta \end{aligned} \tag{3}$$

where $\alpha=[\alpha_1,\ldots,\alpha_n]^T$ and $\beta=[\beta_1,\ldots,\beta_n]^T$ with their respective entries, α_i, β_i as the linear combination coefficients. This is usually called "dual representation" [19].

2.2. KCCA

In KCCA [3], we adopt two nonlinear mappings: $\Phi_x : x \longrightarrow \mathcal{F}_x^{n_x}$ and $\Phi_y : y \longrightarrow \mathcal{F}_y^{n_y}$. Here, n_x and n_y represent dimensions of these two feature spaces $\mathcal{F}_x^{n_x}$ and $\mathcal{F}_y^{n_y}$. By these two mappings, original features x_i and y_i can be mapped as $\Phi_x(x_i)$ and $\Phi_y(y_i)$, where $i=1,2,\ldots,n$. For basis vector pair $(\omega_{\Phi_x}, \omega_{\Phi_y})$ in the feature space, they have corresponding dual representations as below:

$$\begin{aligned} \omega_{\Phi_x} &= \Phi_x(X)\alpha \\ \omega_{\Phi_y} &= \Phi_y(Y)\beta \end{aligned} \tag{4}$$

where $\Phi_x(X)=\{\Phi_x(x_1),\ldots,\Phi_x(x_n)\}$ and $\Phi_y(Y)=\{\Phi_y(y_1),\ldots,\Phi_y(y_n)\}$. $\Phi_x(X)$ and $\Phi_y(Y)$ can be treated as mapped feature sets of X and Y respectively. The means of $\Phi_x(X)$ and $\Phi_y(Y)$ are both zeros. Then $\alpha=[\alpha_1,\ldots,\alpha_n]^T$ and $\beta=[\beta_1,\ldots,\beta_n]^T$ denote corresponding dual coefficient vectors in feature spaces. Similar to Eq. (1), for KCCA, the objective function to be maximized can be written as