A Deep Covariation Projection for Latent Multiview Representation

Yun-Hao Yuan, Jin Li, Yun Li, Jipeng Qiang, Yi Zhu

***Abstract*—Canonical correlation analysis (CCA) has recently attracted proliferating interests in multiview data learning by finding the relationship between two multidimensional data sets. However, CCA-related methods are mostly dependent on covariance matrices in essence, thus making them incapable of modeling complex nonlinear relationship among different *feature vectors*. To address this problem, we propose an $\mathcal{F}$-deep covariation projection approach dubbed $\mathcal{F}$-DCP for learning latent multiview representation, key idea of which is based on deep covariation matrices instead of classical covariance matrices. The proposed method, which consists of two components: $\mathcal{F}$-deep CCA and alignment module, projects each *feature* rather than *sample* in the original two view spaces into a certain new space via deep neural networks, which is the most prominent difference between our method and conventional deep CCA work, and thus leads to more flexibility and applicability in practice. Furthermore, we present an extension of $\mathcal{F}$-DCP for learning latent representation from more than two views. Extensive experimental results on benchmark datasets demonstrate that our $\mathcal{F}$-DCP method outperforms the state-of-the-art methods in terms of clustering evaluation and classification accuracy.**

***Index Terms*—Multiview representation learning, canonical correlation analysis, deep learning, multiview clustering.**

I. INTRODUCTION

In many machine learning applications, the same objects have multiple representations from different viewpoints. For instance, an image can be depicted by texture and shape features. Web pages can be described by text and hyperlink information. Such data are referred to as multiview data [1]. In each view, the observations are usually represented in a high-dimensional space. High dimensionality significantly increases the difficulty of discriminating objects from different categories and leads to the problem of "curse of dimensionality" [2]. Thus, how to jointly learn latent feature representation effectively and efficiently from high-dimensional multiview data has become a fundamental problem in numerous practical multiview learning tasks [3].

An appealing research direction to tackle this problem is multiview representation learning (MRL) [4], which aims to learn useful latent representation of multiview data to improve the learning performance of downstream tasks. Since multiview data have remarkable advantages such as complementarity and diverse statistical properties, MRL can learn more discriminative representation than single view representation learning for real applications. To date, many MRL methods have been developed based on different motivations; see, e.g., [4]–[7]. Among the most representative is canonical correlation analysis (CCA) [8], which is an effective subspace learning method for seeking the correlation between two multidimensional data sets (views) in an unsupervised fashion. It has found numerous practical applications in, for example, multiview clustering [9], [10], multilabel learning [11], information fusion [12], and understanding of deep networks [13].

The objective of CCA is to find a pair of linear transformations, one for each view, such that the projected two-view data in the latent low-dimensional space are maximally correlated. It is well-known that CCA has a closed-form solution to the optimal linear transformations, which can be obtained by solving a generalized eigenequation [11]. Theoretically, CCA can also be interpreted as a latent variable model [14] from a probabilistic perspective, thus enabling it to be included in a larger probabilistic graphical model. The last few years have seen a number of developments of CCA for latent MRL; see, for instance, regularized CCA [11], sparse CCA [15], [16], $L_{2,1}$-CCA [17], and orthogonal CCA (OCCA) [18].

Despite achieving promising results for MRL, most methods of CCA and its extensions still confront a considerable challenge. Recall that, existing CCA approaches are mostly dependent on the so-called *spectral* methods, which are based on the associated eigenvectors corresponding to the top or bottom eigenvalues of specially constructed data matrices. Such data matrices are, in essence, based on explicit or implicit covariance matrices that depict the linear relationship of different features. This implies that they are not capable of modeling the nonlinear relationship hidden in *feature vectors*[1]. To illustrate this clearly, considering two feature vectors $\mathbf{f}_1 \in \mathbb{R}^n$ and $\mathbf{f}_2 \in \mathbb{R}^n$ with n as the number of samples, their covariance is statistically defined by

$$\begin{aligned}\mathrm{Cov}(\mathbf{f}_1, \mathbf{f}_2) &= \frac{1}{n}\left(\mathbf{f}_1 - \frac{1}{n}\mathbf{1}^T\mathbf{f}_1\mathbf{1}\right)^T\left(\mathbf{f}_2 - \frac{1}{n}\mathbf{1}^T\mathbf{f}_2\mathbf{1}\right) \\ &= \left\langle \frac{1}{\sqrt{n}}\left(\mathbf{f}_1 - \frac{1}{n}\mathbf{1}^T\mathbf{f}_1\mathbf{1}\right),\ \frac{1}{\sqrt{n}}\left(\mathbf{f}_2 - \frac{1}{n}\mathbf{1}^T\mathbf{f}_2\mathbf{1}\right)\right\rangle,\end{aligned} \tag{1}$$

where $\mathrm{Cov}(\cdot,\cdot)$ is the covariance operator, $\langle\cdot,\cdot\rangle$ denotes the inner product of two vectors, and $\mathbf{1} \in \mathbb{R}^n$ is a column vector with all entries as 1. It is clear that the covariance defined above is the inner product of two scaled vectors, which means

Yun-Hao Yuan is with the School of Information Engineering, Yangzhou University, Yangzhou 225127, China and he is also with the School of Computer Science, Fudan University, Shanghai 200433, China (e-mail: yhyuan@yzu.edu.cn).

Jin Li, Yun Li, Jipeng Qiang, and Yi Zhu are with the School of Information Engineering, Yangzhou University, Yangzhou 225127, China (e-mail: cvjinli@outlook.com; liyun@yzu.edu.cn; jpqiang@yzu.edu.cn; z8d1177@126.com).

[1]In this paper, we call the row in $\mathbf{X} \in \mathbb{R}^{d\times n}$ *feature vector* (or *feature* without confusion), while the column *sample*, where d is the dimension of samples and n is number of samples.