J. Vis. Commun. Image R. 56 (2018) 256–264

Contents lists available at ScienceDirect

J. Vis. Commun. Image R.

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

Unsupervised multi-view feature extraction with dynamic graph learning

Dan Shi [a], Lei Zhu [a,*], Zhiyong Cheng [c], Zhihui Li [b], Huaxiang Zhang [a]

[a] School of Information Science and Engineering, Shandong Normal University, China
[b] School of Computer Science and Engineering, University of New South Wales, Australia
[c] School of Computing, National University of Singapore, Singapore

ARTICLE INFO

Article history:
Received 16 May 2018
Revised 7 September 2018
Accepted 25 September 2018
Available online 26 September 2018

Keywords:
Multi-view feature extraction
Intrinsic sample relations
Dynamic graph learning

ABSTRACT

Graph-based multi-view feature extraction has attracted much attention in literature. However, conventional solutions generally rely on a manually defined affinity graph matrix, which is hard to capture the intrinsic sample relations in multiple views. In addition, the graph construction and feature extraction are separated into two independent processes which may result in sub-optimal results. Furthermore, the raw data may contain adverse noises that reduces the reliability of the affinity matrix. In this paper, we propose a novel Unsupervised Multi-view Feature Extraction with Dynamic Graph Learning (UMFE-DGL) to solve these limitations. We devise a unified learning framework which simultaneously performs dynamic graph learning and the feature extraction. Dynamic graph learning adaptively captures the intrinsic multiple view-specific relations of samples. Feature extraction learns the projection matrix that could accordingly preserve the dynamically adjusted sample relations modelled by graph into the low-dimensional features. Experimental results on several public datasets demonstrate the superior performance of the proposed approach, compared with state-of-the-art techniques.

1. Introduction

High-dimensional features have been widely used for complex data representation in many research fields such as multimedia computing, data mining, pattern recognition and machine learning. However, high-dimensional features lead to the problem of "curse of dimensionality" and bring great computation pressure on the machine learning models. Dimensionality reduction can mitigate the problem by identifying the low-dimensional latent subspace that could preserve the data similarities in original high-dimensional space. It is generally achieved by two common paradigms: feature selection and feature extraction. Feature selection chooses a subset of the original features as low-dimensional representations by dropping out irrelevant and noisy features, while feature extraction learns a specific transformation matrix to generate projected dimensions that can still preserve the inherent data characteristics. According to the dependence on semantic labels, feature extraction can be further divided into two families: unsupervised and supervised feature extraction. In this paper, we mainly focus on the learning paradigm of unsupervised feature extraction technique.

Unsupervised feature extraction generates low-dimensional features without considering any explicit semantic labels. Due to its desirable performance, many works have been developed following this paradigm in the past few decades. Multidimensional Scaling (MDS) [1] finds an embedding subspace that preserves the interpoint distances during dimensionality reduction. Principal Component Analysis (PCA) [2] preserves the statistic variance measured in the high-dimensional input space into low-dimensional embedding of data points. Isometric Feature Mapping (Isomap) [3] extends MDS by incorporating the geodesic distances modeled by a weighted graph. Locally Linear Embedding (LLE) [4] maintains the local linearity of the sample during dimensionality reduction. Locality Preserving Projection (LPP) [5] learns linear projection maps by solving a variational problem that optimally preserves the neighborhood structure of data. The main limitation of these methods is that they can only deal with the feature extraction problem on single-view data.

In contrast, real-world data is actually complex and multiple features should be extracted to more accurately describe data contents. In visual domain, an image is generally described by diverse descriptors, such as GIST [6], SIFT [7] and HOG [8]. In audio domain, an audio clip is usually represented by several audio features, such as MFCC [9], LPC [10] and PLP [11]. Apparently, multi-view data can capture the inherent data correlations from

* Corresponding author.
E-mail address: leizhu0608@gmail.com (L. Zhu).

https://doi.org/10.1016/j.jvcir.2018.09.019
1047-3203/

different aspects with more accuracy and comprehensiveness [12–17].

Multi-view feature extraction is proposed to exploit the relevance and complementarity of multi-view data. Typical examples include Marginal Fisher Analysis (MFA) [18], Multi-View Spectral Embedding (MSE) [19] and Multi-View Locally Linear Embedding (MLLE) [20]. These methods are generally based on graph theory. In them, several fixed graphs are first constructed to represent data similarity in multiple views separately. Then, these graphs are integrated into a unified one, based on which the ultimate feature extraction is performed. Even though these methods achieve impressive performance, they still suffer from several drawbacks: (1) The graph construction and feature extraction are separated into two independent processes, which tends to lead to suboptimal results. (2) They learn the extracted features with the fixed affinity graph matrix. Real-world data always contain noises that are harmful to the quality of affinity graph. Thus, the subsequent feature extraction performance may be impaired accordingly. (3) They suffer from the out-of-sample problem. They cannot process the new data points that are not included in the training set.

To solve these problems, in this paper, we propose a novel unsupervised multi-view feature extraction method with dynamic graph learning. The main contributions of this paper are summarized as follows:

- We devise a joint unsupervised multi-view feature extraction learning framework that learns a feature extraction matrix and an dynamic graph simultaneously. This framework enables the feature extraction matrix to possess satisfactory projection ability that can preserve the modeled data correlations. Meanwhile, the dynamic graph can adaptively model the correlational relationships between multi-view data. To the best of our knowledge, there is still no similar work.
- An effective optimization solution guaranteed with desirable convergence is proposed to iteratively learn the optimal view combination weights, dynamic graph structure and feature extraction matrix. It can reach to optimal solution after finite iterations, which has conspicuous advantage in unsupervised multi-view feature extraction.
- Extensive experiments on public multi-view datasets demonstrate the proposed method can achieve state-of-the-art performance, and also validate the desirable advantage of dynamic graph learning on multi-view feature extraction.

The rest of this paper is organized as follows. In Section 2, we briefly review the related work. In Section 3, we detail the proposed approach. In Section 4, we show the theoretical analysis about UMFE-DGL. In Section 5, we present the experiments. Finally, we conclude the paper in Section 6.

2. Related work

In this section, we briefly review the related research on unsupervised feature extraction and multi-view learning.

2.1. Unsupervised feature extraction

Unsupervised feature extraction projects high-dimensional data into the low-dimensional subspace with similarity preserving. It has been widely utilized in many fields such as pattern recognition and machine learning. Various methods have been developed in these research areas. They roughly include linear and nonlinear learning paradigms. Principal Component Analysis (PCA) [2] is a typical linear unsupervised feature extraction method. The core of PCA is to map high-dimensional data to a low-dimensional space through linear projection, while preserving the perspective of covariance of features. The limitation of PCA is that it cannot ensure the learned subspace to be discriminative. Locally linear embedding (LLE) [4] is a representative non-linear feature extraction method that can maintain the original manifold structure into the reduced data dimensions with locally linear embedding. However, it is sensitive to the number of nearest neighbors, which has a great impact on the feature extraction performance. Projective Unsupervised Flexible Embedding Models with Optimal Graph (PUFE-OG) [21] proposes flexible graph learning to reduce dimensions for image and video processing, but the graph learning relies on a fixed graph that may be unreliable.

2.2. Multi-view learning

In many real world applications, data are often collected from different views since single view data cannot comprehensively express the example [22–25]. Thus, many multi-view learning approaches are proposed and they have benefited for many applications. For instance, [26] develops a multiple social network learning model to predict volunteerism tendency. [27,28] propose multi-source multi-task learning scheme to achieve user interest prediction. [29] introduces a multi-view transfer learning framework to predict image memorability. [30] focuses on popularity prediction of micro-videos by presenting a low-rank multi-view embedding learning framework. [31–36] also consider learning with multiple views to improve the performance.

In feature extraction, multi-view methods are proposed to exploit the complementary and correlation of multi-view features. For example, Multi-View Spectral Embedding (MSE) [19] first builds patches for samples on different views, and then obtains the low-dimensional embedding by the part optimization. Finally, all low-dimensional embeddings from different patches are unified as a integrated one. The major problem of MSE is that it requires all feature matrices to perform matrix decomposition, which will suffer from great computation complexity. Multi-View Locally Linear Embedding (MLLE) [20] preserves the geometric structure of the local patch into the low-dimensional embedding according to the locally linear embedding criterion. Although these methods have good performance, they learn the extracted feature with fixed graph matrices. Besides, the graph construction and feature extraction are separated into two independent processes without any interaction. Thus, the sub-optimal feature extraction performance may be possibly brought. Furthermore, real-world data inevitably contain noises. The quality of the relied affinity graphs may be impaired and thus the feature extraction performance may be degraded. Unsupervised Multiple Views Feature Extraction with Structured Graph (MFESG) [37] learns the feature extraction matrix and the ideal structure graph simultaneously, and assigns a weight factor for each view. This method aims to learn a structured graph for feature extraction. However, it performs the graph learning on a fixed affinity graph matrix, whose quality directly determines the quality of learned structured graph and the ultimate feature extraction performance under this circumstance.

Different from the above methods, in this paper, we directly learn the optimized dynamic graph from raw features without dependence on any pre-constructed graph. Moreover, we carefully consider the different contributions of multi-view features on learning dynamic graph by assigning them differentiated importance weights. To the best of our knowledge, there is still no similar work.

3. Methodology

In this section, we will introduce the proposed UMFE-DGL in detail. First, we give the relevant notations and definitions used