Neural Networks 103 (2018) 1–8

Contents lists available at ScienceDirect

Neural Networks

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

Beyond Low-Rank Representations: Orthogonal clustering basis reconstruction with optimized graph structure for multi-view spectral clustering

Yang Wang [a,b,*], Lin Wu [c]

[a] Dalian University of Technology, Dalian 116024, China
[b] The University of New South Wales, NSW 2052, Australia
[c] The University of Queensland, Queensland 4072, Australia

ARTICLE INFO

Article history:
Received 31 July 2017
Received in revised form 19 December 2017
Accepted 9 March 2018
Available online 20 March 2018

Keywords:
Low-Rank Representation
Multi-view subspace learning
Clustering

ABSTRACT

Low-Rank Representation (LRR) is arguably one of the most powerful paradigms for Multi-view spectral clustering, which elegantly encodes the multi-view local graph/manifold structures into an intrinsic low-rank self-expressive data similarity embedded in high-dimensional space, to yield a better graph partition than their single-view counterparts. In this paper we revisit it with a fundamentally different perspective by discovering LRR as essentially a latent clustered orthogonal projection based representation winged with an optimized local graph structure for spectral clustering; each column of the representation is fundamentally a cluster basis orthogonal to others to indicate its members, which intuitively projects the view-specific feature representation to be the one spanned by all orthogonal basis to characterize the cluster structures. Upon this finding, we propose our technique with the following: (1) We decompose LRR into latent clustered orthogonal representation via low-rank matrix factorization, to encode the more flexible cluster structures than LRR over primal data objects; (2) We convert the problem of LRR into that of simultaneously learning orthogonal clustered representation and optimized local graph structure for each view; (3) The learned orthogonal clustered representations and local graph structures enjoy the same magnitude for multi-view, so that the ideal multi-view consensus can be readily achieved. The experiments over multi-view datasets validate its superiority, especially over recent state-of-the-art LRR models.

1. Introduction

Spectral clustering (Ng, Jordan, & Weiss, 2001), which partitions the data objects via their local graph/manifold structure relying on the Laplacian eigenvalue–eigenvector decomposition, is one fundamental clustering problem. Unlike K-Means clustering (Wu et al., 2008), the data objects within the same group characterize not only the large data similarity but also the similar local graph/manifold structure. With the rapid development of information technology, the data are largely available with the multi-view feature representations (*e.g.*, images can be featured by a color histogram view or a texture view), which naturally paves the way to multi-view spectral clustering. As extensively claimed by the multi-view research (Deng et al., 2015; Li, Nie, Huang, & Huang, 2015; Nie, Cai, & Li, 2017; Xu, Tao, & Xu, 2015), the information encoded by multi-view features describes different properties; thus leveraging the multi-view information can outperform the single-view counterparts. One critical issue on a successful multi-view incorporation implied by the existing work (Gui et al., 2014; Kumar & Daume, 2011; Wang, Lin, Wu, & Zhang, 2017; Wang et al., 2015; Wang, Wu, Lin, & Gao, 2018; Wang, Zhang, Wu, Lin, & Zhao, 2017) lies in how to achieve the multi-view consensus/agreement.

Following such principle, a lot of multi-view clustering methods (Gao, Han, Liu, & Wang, 2013; Gao, Nie, Li, & Huang, 2015) claim that similar data objects should be within the same group across all views. Based on that, the consensus multi-view local manifold structure is further explored with great efforts (Kumar & Daume, 2011; Kumar, Rai, & Daume, 2011; Wang et al., 2016; Xia, Pan, Du, & Yin, 2014) for multi-view spectral clustering. Among all these methods, Low-Rank Representation (LRR) (Liu, Lin, & Yu, 2010) coupled with sparse decomposition based model has been emerged as a substantially elegant solution, due to its strength of exploring their intrinsic low-dimensional manifold structure encoded by the data correlations embedded in high-dimensional

* Corresponding author at: Dalian University of Technology, Dalian 116024, China.
E-mail addresses: yang.wang@dlut.edu.cn (Y. Wang), lin.wu@uq.edu.au (L. Wu).

https://doi.org/10.1016/j.neunet.2018.03.006
0893-6080/

space, while exhibiting strong robustness to feature noise corruptions addressed by sparse noise modeling, hence attracting great attention.

Before proceeding further, some notations that are used throughout the paper are shown below.

1.1. Notations

For Matrix M, the trace of M is denoted as $Tr(M)$; $\|M\|_F = \sqrt{\sum_{i,j} M_{i,j}^2}$ (or $\|\cdot\|_2$ for vector space) denotes the Frobenius norm; $\|M\|_1 (\sum_{i,j} |M_{i,j}|)$ is the ℓ_1 norm, and M^T denotes the transpose of M, and its unclear norm as $\|M\|_*$ (sum of all singular values); $M(i, \cdot)$ and $M(\cdot, i)$ as the ith row and column of M. $M \succeq 0$ means all entries of M are nonnegative. I is the identity matrix with adaptive size. $\mathbf{1}$ indicates the vector of adaptive length with all entries to be 1. $|\cdot|$ indicates the cardinality of the set.

1.2. Motivation: LRR revisited for multi-view spectral clustering

Specifically, the typical LRR model for multi-view spectral clustering stems from the formulation below:

$$\min_{Z, E_i} \|Z\|_* + \lambda \sum_{i \in V} \|E_i\|_1 \quad \text{s.t. } X_i = X_i Z + E_i, \;\; i \in V, \;\; Z \succeq 0, \tag{1}$$

where $X_i \in \mathbb{R}^{d_i \times n}$ is the data representation for the ith view with d_i as its feature dimension, n as the number of data objects identical for each view, λ is the balance parameter, and V is the view set. $Z \in \mathbb{R}^{n \times n}$ is the self-expressive low-rank similarity representation shared by all $|V|$ views, constrained with $\|Z\|_*$ based on $X_i (i \in V)$, which can also be substituted by the other specific dictionaries; $\|E_i\|_1$ is modeled to address the noise-corruption for the ith view-specific feature representation. $Z \succeq 0$ ensures the nonnegativity for all its entries. Based on such optimized low-rank Z, the spectral clustering is finally conducted. One significant limitation of Eq. (1) pointed out by Wang et al. (2016) is that, *only one common Z* is learned to *preserve the flexible local manifold structures for all views*, hence fails to achieve the ideal spectral clustering result. To this end, various low-rank Z_i are learned to preserve the ith view-specific local manifold structures, meanwhile minimize their divergence via an iterative-views-agreement strategy for multi-view consensus, followed by a final spectral clustering stage. Despite its encouraging performance, the following standout limitations are inattentively overlooked for LRR model: **(1)** The low-rank data similarity may not well encode the flexible latent cluster structures over primal view-specific feature space; worse still for the non-ideal local graph construction over such representation for spectral clustering; **(2)** The low-rank data similarities coming from multi-views may not be within the same magnitude, so that the divergence minimization may not achieve the ideal multi-view clustering consensus.

Our new perspective. The above facts motivate us to revisit the low-rank representation Z_i to help $X_i Z_i$ reconstruct X_i below for the ith view

$$\min_{Z_i \in \mathbb{S}} \|X_i - X_i Z_i\|_F^2, \tag{2}$$

where $\mathbb{S}$ denotes the set of $Z_i \in \mathbb{R}^{n \times n}$ with low-rankness *e.g.,* cluster number c far less than d_i; Instead of narrowing Low-Rank Z_i as self-expressive data similarity from the conventional viewpoint, it is essentially seen as a special case of a generalized Low-Rank projection, to map feature representation to a low-dimensional space to reconstruct X_i with minimum error. As discussed, the self-expressive similarity projection equipped with LRR models still suffer from the aforementioned non-trivial limitations.

Here we ask a question: Is there a superior low-rank projection Z_i to minimize Eq. (2), meanwhile address the limitations over the existing LRR models. Our answer to this question is positive. Specifically, we propose to consider Z_i as a latent clustered orthogonal projection, via $Z_i = U_i U_i^T$, where

1. **Clustered orthogonal projection**: $U_i \in \mathbb{R}^{n \times c}$, where each column indicates one cluster to characterize its belonging data objects. Compared with LRR over original feature space, the latent factor U_i can better preserve the flexible latent cluster structure.
2. **Feature reconstruction with cluster basis**: Instead of low-rank data similarity, Z_i essentially serves as a mapping to reconstruct the view-specific features via the column of U_i to encode the latent cluster structures.
3. **Rethinking** $X_i Z_i$: We revisit the intuition of $X_i Z_i$ via $(X_i U_i) U_i^T$ throughout two stages, remind that $X_i \in \mathbb{R}^{d_i \times n}$ where
   - $X_i U_i$ is performed to obtain the new projection value for all d_i features over c orthogonal columns of U_i;
   - $X_i U_i U_i^T$ is subsequently the projected representation for all d_i features spanned by c clustered orthogonal column basis of U_i.
4. **Same magnitude for multi-view consensus**: All $U_i (i \in V)$ enjoy the same magnitude due to their orthonormal columns. Hence, the feasible divergence minimization will facilitate the multi-view consensus.

Before shedding light on our technique, we review the typical related work for multi-view spectral clustering.

1.3. Prior arts

The prior arts can be classified as per the strategy at which the multi-view fusion takes place for spectral clustering.

The most straightforward method goes to the *Early fusion* (Huang, Liu, Zhang, & Metaxas, 2010) by concatenating the multi-view feature vectors with equal or varied weights into an unified one, followed by the spectral clustering over such unified space. However, such method ignores the statistical property belonging to an individual view. *Late fusion* (Greene & Cunningham, 2009) may address the limitation to some extents by aggregating the spectral clustering result from each individual view, which follows the assumption that all views are independent to each other. Such assumption is not effective for multi-view spectral clustering as they assume the views to be dependent so that the multi-view consensus information can be exploited for promising performance.

Canonical Correlation Analysis (CCA) is applied for multi-view spectral clustering (Chaudhuri, Kakade, Livescu, & Sridharan, 2009) by learning a common low-dimensional representations for all views, upon which the spectral clustering is performed. One salient drawback lies in the failure of preserving the flexible local manifold structures for different views via such common subspace. *Co-training* based model (Kumar & Daume, 2011) learned the Laplacian eigenmap for each view over its projected data representation throughout the Laplacian eigenmaps from other views, such process repeated till the convergence, the final similarity are then aggregated for spectral clustering. A similar method (Kumar et al., 2011) is also proposed to coordinate multi-view Laplacian eigenmaps consensus for spectral clustering. Despite their effectiveness, they have to follow the scenario of noise free for the feature representations. Unfortunately, it cannot be met in practice. The Low-Rank Representation and sparse decomposition models (Wang et al., 2016; Xia et al., 2014) well tackle the problem, meanwhile exhibits the robustness to feature noise corruptions. However, they still suffer from the aforementioned limitations. To this end, we make the following orthogonal contributions to typical LRR model for multi-view spectral clustering.