Neurocomputing 249 (2017) 106–116

Contents lists available at ScienceDirect

Neurocomputing

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

Integrating feature and graph learning with low-rank representation

Chong Peng[a,b,*], Zhao Kang[b], Qiang Cheng[b]

[a] Department of Applied Mathematics, Yuncheng University, Yuncheng, Shanxi Province, 044000, China
[b] Department of Computer Science, Southern Illinois University Carbondale, IL, 62901, USA

ARTICLE INFO

Article history:
Received 3 August 2016
Revised 24 December 2016
Accepted 27 March 2017
Available online 5 April 2017

Communicated by Deng Cai

Keywords:
Subspace clustering
Feature selection
Manifold learning
Robust

ABSTRACT

We propose a new subspace clustering method that integrates feature and manifold learning while learning a low-rank representation of the data in a single model. This new model seeks a low-rank representation of the data using only the most relevant features in both linear and nonlinear spaces, which helps reveal more accurate data relationships in both linear and nonlinear spaces, because data relationships can be less afflicted by irrelevant features. Moreover, the graph Laplacian is updated according to the learning process, which essentially differs from existing nonlinear subspace clustering methods that require constructing a graph Laplacian as an independent preprocessing step. Thus the learning processes of features and manifold mutually enhance each other and lead to powerful data representations. Extensive experimental results confirm the effectiveness of the proposed method.

1. Introduction

It has been increasingly common to use high-dimensional datasets, such as images, videos, text, and more others, in machine learning and data mining tasks. High-dimensional data often lie on a union of multiple low-dimensional structures rather than uniformly distributed [1] and therefore it is helpful to preserve and reveal latent structures of such data by recovering low-dimensional subspaces. For example, in Fig. 1, rather than uniformly distributed in the $\mathcal{R}^3$ space, the data points lie on a union of two lines and two planes, which forms a lower-dimensional structure.

To find such a low-dimensional structure, one usually needs to cluster data points into different groups such that each group can be fitted with a subspace. Recovering the underlying subspaces introduces the problem of subspace segmentation, which is formally defined as follows:

Definition 1. (Subspace segmentation [2,3]) Given a set of sample vectors $X = [X_1, \ldots, X_k] = [x_1, \ldots, x_n] \in \mathcal{R}^{d\times n}$ drawn from a union of k subspaces $\{\mathcal{S}_i\}_{i=1}^k$, where X_i is a collection of n_i samples drawn from the subspace $\mathcal{S}_i$, with $n = \sum_{i=1}^k n_i$, the task of subspace segmentation is to segment the samples according to the underlying subspaces they are drawn from.

Subspace segmentation has been widely studied due to its numerous applications in computer vision, such as face recognition [4,5], motion segmentation [6,7], and image segmentation [8,9]. Subspace segmentation methods can be mainly divided into four categories: algebraic methods [10], statistical methods [9], iterative methods [11], and spectral clustering based methods [1,12,13]; see [2] for reviews. Recently, spectral clustering based subspace clustering methods have shown success in subspace segmentation. Among them, low-rank representation (LRR) [12] and sparse subspace clustering (SSC) [1] are state-of-the-art methods. These methods construct data relationships using global data information, which seek a representation coefficient matrix of the data by finding a linear representation for each example with respect to the collection of all data points. Then they build a sample affinity matrix based on the representation, followed by spectral clustering [14] to segment the data points. Both LRR and SSC seek a linear representation of the data, however they require different structures of the low-dimensional representation, where LRR requires it to be low-rank while SSC requires it to be sparse, respectively. Recently, some new methods [13,15] marry the advantages from LRR and SSC and imposes simultaneously low-rank and sparse structure on the representation coefficients.

Learning low-rank and sparse models have been well studied [16–18]. However, LRR and SSC model data relationships in the original space, which only consider linear structures of the data in Euclidean space. This approach takes no consideration of nonlinear structures of the data in nonlinear space, which is usually important for clustering tasks. To address this problem, recent studies consider nonlinear structures of the data in a kernel feature space or on a manifold. For example, kernel sparse subspace clustering (KSSC) [19], kernel LRR (RKLRR) [20] find sparse or low-rank representation of the data in kernel feature space; kernelized LRR

* Corresponding author.
E-mail addresses: pchong@siu.edu (C. Peng), zhao.kang@siu.edu (Z. Kang), qcheng@cs.siu.edu (Q. Cheng).

http://dx.doi.org/10.1016/j.neucom.2017.03.071
0925-2312/

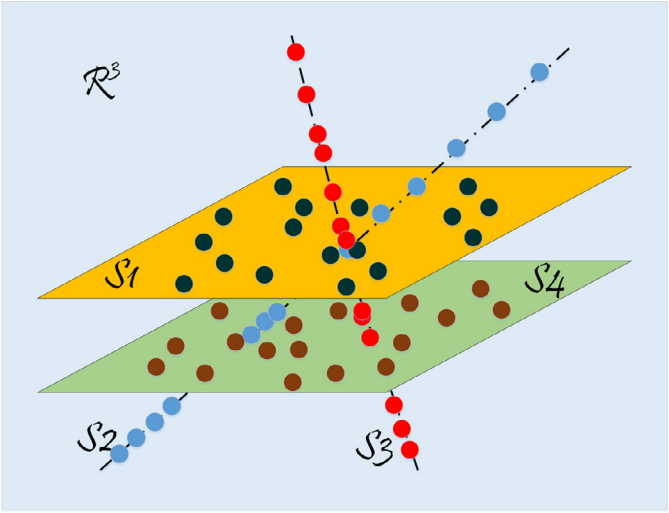

Fig. 1. A set of data points in $\mathcal{R}^3$ are drawn from a union of four subspaces, i.e., two planes and two lines.

on Grassman manifold (KGLRR) [21] finds low-rank representation of the data on manifold. These methods are based on LRR and SSC models, while enhancing the capabilities of recovering nonlinear data relationships by nonlinear techniques. However, the above-mentioned methods, including the linear and nonlinear models, may be problematic by finding the representation coefficients or constructing kernel matrix or graph Laplacian using all features of the data, because it is well known that noisy, redundant, or irrelevant features often exist in high-dimensional data [22] and they may degrade the learning performance on such data. In fact, feature selection has been proven important and thus is often applied as a pre-processing step for learning high-dimensional data [23,24]. For unsupervised learning, there is often a "chicken-and-egg" problem for processing feature selection and clustering, because unsupervised feature selection is often transformed into a supervised one by the existing unsupervised feature selection methods, such as multi-class feature selection (MCFS) [25], which requires clustering for preprocessing. However, this procedure does not meet the demand of finding important features as input for subspace clustering. Moreover, it has been demonstrated that it is inappropriate for some problems to involve a two-step processing procedure [26,27]. To avoid such "chicken-and-egg" situation as well as the potential problem in a two-step learning procedure, feature selection embedded subspace clustering (FSC) [28] integrates feature selection with subspace recovery in a single, seamlessly integrated framework for enhanced performance. In fact, integrating multi-tasks in a single model has been shown successful and important. For example, [29] reduces the dimension of input data with an embedded projection; [30] simultaneously performs semi-NMF and PCA to identify a partition of the data by seeking an optimal subspace of multi-dimensional variables, which provides simpler and more interpretable solutions using low-dimensional representation. FSC learns important features and use these features to build data relationships by finding a low-rank representation of the data. However, it only considers recovering structures of the data with a subspace of features in the Euclidean space, and thus the built representation coefficients only capture linear structures of the data. Therefore, FSC is unable to recovery nonlinear structures of the data. To simultaneously resolve the limitations and marry the advantages of FSC and nonlinear subspace clustering methods, this paper proposes to simultaneously perform feature selection and subspace recovery in the original instance space as well as on a manifold, which enhances the capability of constructing both linear and nonlinear structures of the data using most relevant features to essentially improve the clustering performance. In this case, the graph Laplacian is iteratively adapted according to the feature selection and learning process, which renders cleaner graph Laplacian to be learned. It is noted that this strategy overcomes a general drawback of the nonlinear subspace clustering methods, which constructs the kernel matrix or graph Laplacian as a preprocessing step because they are built independently before processing the data.

We summarize the main contributions of this paper as follows:

- A new subspace clustering method is proposed which seamlessly incorporates feature learning to alleviate the adverse effect of irrelevant or less important features when finding the representation coefficients for the data;
- The new subspace clustering model incorporates manifold learning, which enhances the capability of exploiting nonlinear structures of the data. It constructs the graph based on the most relevant features obtained during the learning process, and thus the graph is less afflicted with irrelevant and grossly corrupted features and is more discriminative. The graph Laplacian is learned iteratively based on the feature learning process, which starkly differs from widely used nonlinear techniques that construct a kernel matrix or graph Laplacian as input from an independent preprocessing step. Therefore, the proposed subspace clustering method assimilates as well as enables feature learning and manifold learning jointly, providing a powerful data representation;
- The integrated model allows us to simultaneously achieve multiple objectives in a mutually enhancing manner, that are essentially important for more representative data representation;
- Our framework allows for efficient optimization, where we observe fast convergence in the experiments;
- Extensive experimental results verify the effectiveness of the proposed model and algorithm and show significant improvement compared to state-of-the-art algorithms.

The rest of this paper is organized as follows. We briefly discuss closely related work in Section 2. Then we present the new subspace clustering model in Section 3. Its optimization and complexity are discussed in Sections 4 and 5, respectively. Extensive experimental results are demonstrated in Section 7. Finally, Section 8 concludes the paper.

2. Related work

Recently, spectral clustering-based subspace clustering methods have drawn increasing attentions due to their effectiveness in real world applications [1,12,31–35]. Among these methods, SSC and LRR have become state-of-the-art due to elegant theories and promising performance in handling clustering tasks. Given a collection of n data points, $X = [x_1, \ldots, x_n] \in \mathcal{R}^{d\times n}$, drawn from k subspaces, where each d-dimensional column vector is a data point, the basic assumption of SSC and LRR is that each data point can be represented by all data points and thus the data is self-expressive and can be modeled as $X = XZ + E$, where $Z \in \mathcal{R}^{n\times n}$ is the coefficient matrix and $E \in \mathcal{R}^{d\times n}$ denotes the fitting error. In particular, SSC requires Z to be sparse while LRR requires Z to have low rank, which leads to the following two models:

$$\text{(SSC)} \quad \min_{Z} \|Z\|_1 + \lambda\|S\|_1 + \gamma\|E\|_F^2$$

$$s.t. \quad X = XZ + E + S, \mathbf{1}_n^T Z = \mathbf{1}_n, \text{diag}(Z) = 0, \tag{1}$$

and

$$\text{(LRR)} \quad \min_{Z} \|Z\|_* + \lambda\|E\|_{2,1} \quad s.t. \quad X = XZ + E, \tag{2}$$

where $\lambda > 0$ and $\gamma > 0$ are two balancing parameters, S is the sparse corruptions, $\mathbf{1}_n$ is a n-dimensional column vector of 1s,