Neurocomputing 000 (2018) 1–10

Contents lists available at ScienceDirect

Neurocomputing

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

Discriminative and coherent subspace clustering

Huazhu Chen, Weiwei Wang*, Xiangchu Feng, Ruiqiang He

School of Mathematics and Statistics, Xidian University, Xi'an 710071, China

ARTICLE INFO

Article history:
Received 9 June 2017
Revised 6 November 2017
Accepted 2 January 2018
Available online xxx

Communicated by Dr. Haijun Zhang

Keywords:
Subspace clustering
Discrimination
Coherence
Affinity
Label

ABSTRACT

The ubiquitous large, complex and high dimensional datasets in computer vision and machine learning generate the problem of subspace clustering, which aims to partition the data into several low dimensional subspaces. Most state-of-the-art methods divide the problem into two stages: first learn the affinity from the data and then infer the cluster labels based on the affinity. The Structured Sparse Subspace Clustering (SSSC) model combines the affinity learning and the label inferring into one unified framework and empirically outperforms the two-stage methods. However, the SSSC method does not fully utilize the affinity and the labels to guide each other. In this work, we present a new regularity which combines the labels and the affinity to enforce the coherence of the affinity for data points from the same cluster and the discrimination of the labels for data points from different clusters. Based on this, we give a new unified optimization framework for subspace clustering. It enforces the coherence and discrimination of the affinity matrix as well as the labels, thus we call it Discriminative and Coherent Subspace Clustering (DCSC). Extended experiments on commonly used datasets demonstrate that our method performs better than some two stage state-of-the-art methods and the unified method SSSC in revealing the subspace structure of high-dimensional data.

1. Introduction

In the past few years, technology advances have made data collection easier and faster, resulting in large, multimodal and high-dimensional datasets. How to effectively compress, store, transmit and process massive amounts of such complex high-dimensional data has become a necessary and urgent task. Many existing methods [1] have exploited the observation that the high-dimensional data usually lie in a union of several low-dimensional subspaces or affine spaces. For instance, the face images of a subject obtained under a wide variety of lighting conditions can be accurately approximated with a 9-dimensional linear subspace [2]. This has motivated the problem of subspace clustering, which aims to partition the data points into several low dimensional subspaces and has found numerous applications in computer vision (e.g., image segmentation [3], motion segmentation [4] and face clustering [5]), image processing (e.g., image representation and compression [6]) and systems theory (e.g., hybrid system identification [7]).

Among the existing subspace clustering methods [8–44], the spectral clustering based methods [23–44] are becoming more popular because they are easy to be implemented, and insensitive to initialization and data corruptions. Most spectral clustering based methods divide the problem into two separate stages. First an affinity matrix is learned from the data by using so called self-representation such as Sparse Subspace Clustering (SSC) [28,29], Low-rank Representation (LRR) [30] and some hybrid representation based on SSC or LRR. Then the labels are learned by a spectral clustering method such as Ncut [45]. Although the two-stage methods succeed in many applications, they have a major disadvantage: the relationship between the affinity matrix and the labels of the data is not fully exploited, thus they cannot guarantee an overall optimal performance.

By combining the two stages into one unified framework, the Structured Sparse Subspace Clustering (SSSC) [44] has shown that the overall performance of subspace clustering can be greatly improved. Actually, SSSC uses the self-representation coefficients and the labels to guide each other interactively so that both the affinity and the labels have some advantageous properties. Specifically, it uses the labels to enforce the affinity for data points from different clusters be sparse. Such a property of the affinity is called cluster discrimination property. On the other hand, the self-representation is used to guide the label inferring so that the data points from the same cluster could have same labels. We call this property of the labels coherence property.

Although this unified framework outperforms the two-stage methods, it has some shortcomings. It only enforces the sparseness/discrimination of the affinity matrix for data points from

* Corresponding author.
E-mail addresses: chenhuazhu2001@126.com (H. Chen), wwwang@mail.xidian.edu.cn (W. Wang), xcfeng@mail.xidian.edu.cn (X. Feng), ruiqianghe@sina.com (R. He).

https://doi.org/10.1016/j.neucom.2018.01.006
0925-2312/

different clusters and the coherence of labels for data points from the same cluster. It does not consider the coherence of the affinity for data points from the same cluster or the discrimination of the labels from different clusters. In all, the coupling of the affinity and the labels is not fully exploited.

In this work, we present a new regularity which combines the labels and the affinity to enforce coherence of the affinity and discrimination of labels. We combine it with the structure sparse regularity in SSSC to give a new unified optimization framework for subspace clustering. The main contributions of this work can be summarized as follows:

We present a label-guided regularity to ensure the coherence of the affinity for data points from the same cluster and the discrimination of the labels for data points from different clusters. By combining the label-guided regularity with the structure sparse regularity in SSSC [44], we give a new unified optimization framework for subspace clustering. It enforces the coherence and discrimination of the affinity matrix as well as the labels, thus we call it Discriminative and Coherent Subspace Clustering (DCSC). It can better recover the subspace structure underlying high dimensional datasets and provide more exact clustering results.

Experiments on several commonly used datasets show that our method outperforms other state-of-the-art subspace clustering methods including SSC [28,29], LRR [30], LRSC[32], LSR [33], CASS[34], TSC [39], NSN [41], SSSC[44], LatLRR [46], BDSSC [47], BDLRR[47] and OMP [48].

2. Related works

Let $X = [\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N] \in R^{n\times N}$ be a set of N (sufficiently many) sample points, with each column $\mathbf{x}_i$ being an n-dimensional feature vector, drawn from a union of K subspaces $\{S_c\}_{c=1}^K$ of unknown dimensions $\{r_c\}_{c=1}^K$, respectively. Subspace clustering aims to segment the data into the underlying subspace from which they are drawn.

For convenience, we define some notations used in this work before reviewing related works.

For a matrix $Z = (z_{ij}) = (\mathbf{z}_1, \mathbf{z}_2, \ldots, \mathbf{z}_N) \in R^{N\times N}$ with $\mathbf{z}_j$ being the jth column, $\|Z\|_1 = \sum_{i,j}|z_{ij}|$ and $\|Z\|_F = \sqrt{\sum_{i,j}|z_{ij}|^2}$ are respectively the ℓ_1 -norm and the Frobenius norm of the matrix Z. $\|Z\|_*$ is the trace norm, i.e., the sum of the singular values of the matrix Z. $diag(Z) \in R^{N\times N}$ is the diagonal matrix whose diagonal elements are z_{ii} $(i = 1, \ldots, N)$. $Diag(\mathbf{z})$ denotes a diagonal matrix whose ith diagonal element corresponds to the ith entry of the vector $\mathbf{z}$. $\mathbf{1} \in R^N$ denotes the vector of all $\mathbf{1}'s$.

Define the cluster indicator matrix $Q = (q_{ij}) \in R^{N\times K}$ by

$$q_{ij} = \begin{cases} 1, & if \quad \mathbf{x}_i \in S_j \\ 0, & if \quad \mathbf{x}_i \notin S_j \end{cases} \tag{1}$$

Denote the ith row by $Q(i, :)$ and jth column by $Q(:, j)$, then the row $Q(i, :)$ is the cluster label of the data point $\mathbf{x}_i$ and the column $Q(:, j)$ indicates which points belong to cluster S_j. Assume that each data point lies in exactly one subspace or cluster, then each row of Q has only one entry equal to 1, thus a valid cluster indicator matrix Q satisfies $Q\mathbf{1} = \mathbf{1}$. In addition, it is expected that Q has only K different rows due to K subspaces. So Q also satisfies $rank(Q) = K$. We define the collection of cluster indicator matrices as

$$\mathbb{Q} = \{Q \in \{0, 1\}^{N\times K} : Q\mathbf{1} = \mathbf{1} \; and \; rank(Q) = K\} \tag{2}$$

Let $Q^{(j)}$ be an $N \times N_j$ submatrix of $Diag(Q(:, j))$, consisting of the N_j nonzero columns of the $N \times N$ diagonal matrix $Diag(Q(:, j))$. Based on Q, we define $P = (p_{ij})$ by

$$p_{ij} = \frac{1}{2}\|Q(i, :) - Q(j, :)\|_2^2 = \begin{cases} 1, & if \quad Q(i, :) = Q(j, :) \\ 0, & if \quad Q(i, :) \neq Q(j, :) \end{cases} \tag{3}$$

which indicates whether the data points $\mathbf{x}_i$ and $\mathbf{x}_j$ have the same label, and thus whether they are drawn from the same cluster. So we call P the data connection matrix.

One of the main challenges in spectral clustering-based subspace clustering is how to learn a good affinity matrix $A = (A_{ij})$, where A_{ij} measures the similarity between the data points $\mathbf{x}_i$ and $\mathbf{x}_j$. Recently, many works apply the self-representation to learn the affinity. These methods first find a self-representation matrix Z of the data matrix X by solving the following minimization problem:

$$\min_{Z,E} \Omega(Z) + \lambda\Phi(E) \qquad s.t. \; X = XZ + E, Z \in \mathbb{C}, \tag{4}$$

where $\Omega(Z)$ and $\mathbb{C}$ are the regularity and constraint set, which impose some expected properties on Z. λ is a tradeoff parameter. $\Phi(E)$ is a function penalizing the representation error, corruptions or outliers in the data points. $\|E\|_F^2$ is usually used for Gaussian noise and $\|E\|_1$ is used for sparse entry-wise corruptions. The optimal solution Z^* of problem (4) is used to compute the affinity matrix. A commonly used formula is

$$A = (|Z^*| + |Z^{*T}|)/2 \tag{5}$$

which is further input into a spectral clustering algorithm to produce the final clustering result.

The primary difference between different methods lies in the choice of the regularization term of Z. For example, in the Sparse Subspace Clustering (SSC) [28,29], $\|Z\|_1$ is used as a convex surrogate of $\|Z\|_0$ to promote sparsity of Z. In the Low-Rank Represent (LRR) [30] $\|Z\|_*$ is used to seek a jointly low-rank representation of all data. SSC and LRR show empirical success in some high dimensional datasets. However, a large body of works has shown that SSC performs only optimally in representing data with low correlation and it has the instability problem: if the data from the same subspace are highly correlated or clustered, it will only select one of the several related data at random and ignore other correlated data. This makes it not good for grouping correlated data. The LRR aims at finding the lowest rank representations of all data jointly. It can captured the global structures and not sensitive to noise. However, LRR usually leads to dense representation and results in incorrect clustering. Besides, the number of subspaces and their dimensionality may not be small, thus the data matrix may be high-rank or even full-rank in practice. A number of variants of these algorithms have been proposed, including LatLRR [46], Spatial Weighted SSC [49], LatSSC [50], Kernel SSC [51], etc.

While the above methods have been incredibly successful in many applications, their major disadvantage is that the natural relationship between the coefficient matrix and the segmentation of the data is not explicitly captured. The Structured Sparse Subspace Clustering (SSSC) [44] is a unified optimization framework, which learns the label indicator matrix Q and the self-representation Z simultaneously by solving the following problem:

$$\min_{Z,E,Q} \|Z\|_1 + \alpha\|P \odot Z\|_1 + \lambda\Phi(E) \; s.t. \; X = XZ + E,$$
$$diag(Z) = 0, Q \in \mathbb{Q}, \tag{6}$$

where the operator $\odot$ indicates the Hadamard product (i.e., element-wise product). $\alpha > 0$ and $\lambda > 0$ are tradeoff parameters. The unified framework SSSC empirically outperforms the two-stage methods.

3. Discriminative and coherent subspace clustering: a unified framework

3.1. Motivation

Ideally, the affinity and the labels should be coherent within clusters and discriminative between clusters. Specifically, for the

Please cite this article as: H. Chen et al., Discriminative and coherent subspace clustering, Neurocomputing (2018), https://doi.org/10.1016/j.neucom.2018.01.006