Neurocomputing 135 (2014) 229–239

Contents lists available at ScienceDirect

Neurocomputing

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

Spectral clustering of high-dimensional data exploiting sparse representation vectors

Sen Wu [a], Xiaodong Feng [a,c,*], Wenjun Zhou [b]

[a] Dongling School of Economics and Management, University of Science and Technology Beijing, Beijing 100083, China
[b] Statistics, Operations, and Management Science, University of Tennessee, Knoxville, TN 37996, USA
[c] Department of Computer Science and Engineering, University of Minnesota Twin Cities, Minneapolis, MN 55455, USA

ARTICLE INFO

Article history:
Received 26 June 2013
Received in revised form
12 December 2013
Accepted 19 December 2013
Communicated by D. Tao
Available online 10 January 2014

Keywords:
Spectral clustering
High-dimensional data
Weight matrix
Sparse representation

ABSTRACT

Clustering high-dimensional data has been a challenging problem in data mining and machining learning. Spectral clustering via sparse representation has been proposed for clustering high-dimensional data. A critical stepin spectral clustering is to effectively construct a weight matrix by assessing the proximity between each pair of objects. While sparse representation has proved its effectiveness for compressing high-dimensional signals, existing spectral clustering algorithms based on sparse representation use individual sparse coefficients directly. However, exploiting complete sparse representation vectors is expected to reflect more truthful similarity among data objects, since more contextual information is being considered. The intuition is that sparse representation vectors corresponding to two similar objects are expected to be similar, while those of two dissimilar objects are dissimilar. In particular, we propose two weight matrix constructions for spectral clustering based on the similarity of the sparse representation vectors. Experimental results on several real-world, high-dimensional datasets demonstrate that spectral clustering based on the proposed weight matrices outperforms existing spectral clustering algorithms, which use sparse coefficients directly.

1. Introduction

As an important task in data mining, cluster analysis aims at partitioning data objects into several meaningful subsets, called clusters, such that data objects are similar to those in the same cluster and dissimilar to those in different clusters. With advances in database technology and real-world need of informed decisions, datasets to be analyzed are getting bigger-with many more data records and attributes. Examples of high-dimensional datasets include document data [1], user ratings data [2], multimedia data [3], financial time series data [4], gene expression data [5], and so on. Due to the "curse of dimensionality" [6], clustering high-dimensional data has been a challenging task, and therefore, attracts much attention in data mining and related research domains [7].

Spectral clustering with sparse representation has been found to be effective for clustering high-dimensional data. Spectral clustering [8] is based on the spectral graph model, which is equivalent to graph min-cut problem based on a graph structure constructed from the object space. It is powerful and stable for high-dimensional data clustering [9], and is considered superior to traditional clustering algorithms for high-dimensional data clustering due to its deterministic and polynomial-time solution [8]. Nonetheless, the effectiveness of spectral clustering mainly depends on the input weights between each pair of data objects. Thus, it is vital to construct a weight matrix that faithfully reflects the similarity information among objects. Traditional simple weight construction, such as ε-ball neighborhood, k-nearest neighbors, inverse Euclidean distance [10,11] and Gaussian RBF [9], is based on the Euclidean distance in the original data space, thus not suitable for high-dimensional data due to the "curse of dimensionality" in the original object space. However, sparse representation, coming from compressed sensing [12], proves to be an extremely powerful tool for acquiring, representing, and compressing high-dimensional data by representing each object approximately as a sparse linear combination of other objects. Finding sparse representations transforms the object space into a new sparse space.

Since sparse coefficients represent the contribution of each object to the reconstruction of other objects, existing spectral clustering methods based on sparse representation [13] use these sparse coefficients directly to build the weight matrix. Using the isolated coefficients individually warrants that only local information is utilized. However, we assert that exploiting more contextual information from the whole coefficient vectors promises

* Corresponding author at: Dongling School of Economics and Management, University of Science and Technology Beijing, Beijing 100083, China.
E-mail address: fxdong88@gmail.com (X. Feng).

0925-2312/$ - see front matter
http://dx.doi.org/10.1016/j.neucom.2013.12.027

better assessment of similarity among data objects. Intuitively, our assumption is that the sparse representation vectors corresponding to two similar objects should be similar, since they can be reconstructed in a similar fashion using other data objects.

Therefore, in this paper, we present a study of exploiting contextual information from sparse representation vectors to construct weight matrices for spectral clustering of high-dimensional data. More specifically, we firstly convert each high-dimensional data object into a vector of sparse coefficients, according to sparse representation theories. Then, the proximity of any two data objects is assessed according to the similarity between their sparse representation vectors. We propose two different weight matrix construction approaches: one is based on the consistency of directions, and the other is based on the consistency of magnitude. We show differences and connections between these two different approaches. Finally, spectral clustering is run on the weight matrices constructed. Extensive experiments on several real-world, high-dimensional datasets show that weights exploiting the contextual information from the sparse representation vectors work better than existing solutions, which only utilize individual sparse coefficients, by a variety of clustering performance metrics.

The main contributions of this paper can be summarized as follows. First of all, we recognize the importance of utilizing contextual information for assessing the similarity between data objects. More specifically, in the context of weight matrix construction for spectral clustering, we find that the sparse representation vectors, compared with individual sparse coefficients, contain more details and stronger evidence of similarity between data objects. In addition, we propose two concrete ways to implement the weight matrix construction utilizing sparse representation vectors. Considering the direction of weight contribution, we examine the consistency of the signs for coefficients in the sparse representation vectors. Considering the magnitude of weight contribution, we evaluate the similarity of the sparse representation vectors using the cosine measure. Finally, we validate the usefulness of the proposed approaches with real-world, high-dimensional datasets, showing that they are both better than existing methods, and work better than each other in different scenarios.

The rest of this paper is organized as follows. Related work and some preliminary details are presented in Section 2. In Section 3, we present the two different approaches to construct the weight matrix using sparse representation vectors, and describe the algorithms in detail. In Section 4, we present experiments on several real-world, high-dimensional datasets and evaluate the clustering performances. Finally, we conclude the work in Section 5.

2. Related work and preliminaries

In this section, we first review a few typical techniques for analyzing high-dimensional data. Then we focus on a brief review of the formulation and derivation of sparse representations. Finally, we focus on clustering methods to utilize sparse representations for high-dimensional data.

2.1. Techniques for high-dimensional data

There are many techniques to deal with high-dimensional signals in the literature. Popular techniques include nonnegative matrix factorization, manifold learning, compressed sensing, and combination in between.

Nonnegative matrix factorization (NMF) is a powerful dimensionality reduction technique. It has been widely applied to image processing and pattern recognition [14]. The basic idea is to approximate a non-negative matrix by the product of two non-negative, low-rank factor matrices. It was first proposed by Paatero and Tapper [15], and has attracted much attention in the research community since then. Research on NMF can be generally categorized into three groups. The first group focuses on assessing the consistency between the original matrix and the approximate matrix, using Kullback–Leibler divergence [14], Euclidean distance [16], earth mover's distance [17], Manhattan distance [18], and so on. The second group of research tries to find the optimal solution efficiently and developing scalable NMF algorithms for large-scale datasets. For example, fast Newton-type methods [19], online NMF with robust stochastic approximation [20], robust near-separable NMF using linear optimization [21], and large scale graph regularized NMF [22]. Finally, the third group of research is to improve the performance of NMF under constrains, or exploiting more information from data. These techniques include sparseness constrained NMF [23], convex model for NMF using $l_{1,\infty}$ regularization [24], discriminant NMF [25], graph regularized NMF [26], manifold regularized discriminative NMF [27], and constrained NMF incorporating the label information [28]. In particular, it has been proved that an extended version of NMF is equivalent to kernel K-means and Laplacian-based spectral clustering [29].

Manifold learning is another popular technique to process high-dimensional data, assuming that the data distribution is supported on a low-dimensional sub-manifold [30]. The key idea is that the locality structure of a high-dimensional dataset should be preserved in a low-dimensional space after dimension reduction, which is exploited as a regularization term [31–33] or constraint [34,35] to be added to the original problem formulation. It has been widely used in computer vision applications, such as image classification [36,37], semi-supervised multiview distance metric learning [38], human action recognition [39], and complex object correspondence construction [40].

Besides the two approach reviewed above, sparse representation, originated from compressed sensing, has also attracted a great deal of attention. It proves to be an extremely powerful tool for acquiring, representing, and compressing high-dimensional data. Due to its high relevance to the discussions in this paper, we provide a more detailed overview of spares representation theories in the following subsection.

2.2. A brief review of sparse representation

Given a sufficient high-dimensional training dataset $X=(\boldsymbol{x}_1,\boldsymbol{x}_2,\ldots,\boldsymbol{x}_n)\in\mathbb{R}^{m\times n}$, where $\boldsymbol{x}_i=(x_{i1},x_{i2},\ldots,x_{im})^T\in\mathbb{R}^m$ is a column vector representing the ith object. Research on manifold learning [30] has shown that any test data point $\boldsymbol{y}$ lies on a lower-dimensional manifold, which can be approximately represented by a linear combination of the training data:

$$\boldsymbol{y}=\alpha_1\boldsymbol{x}_1+\cdots+\alpha_i\boldsymbol{x}_i+\cdots+\alpha_n\boldsymbol{x}_n=X\boldsymbol{\alpha}\in\mathbb{R}^m, \tag{1}$$

where $\boldsymbol{\alpha}=(\alpha_1,\alpha_2,\ldots,\alpha_n)^T$ represents the vector of coefficients that need to be determined.

Typically, the number of training objects is much larger than the number of attributes (i.e., $n\gg m$), then Eq. (1) is undetermined, and its solution is not unique.

If we add the constraint that the best solution of $\boldsymbol{\alpha}$ in Eq. (1) should be as sparse as possible, which means that the number of non-zero elements is minimized, then the solution becomes unique. Such a sparse representation can be obtained by solving the following optimization problem:

$$\boldsymbol{\alpha}^*=\arg\min_{\boldsymbol{\alpha}}\|\boldsymbol{\alpha}\|_0 \quad \text{subject to } \boldsymbol{y}=X\boldsymbol{\alpha}, \tag{2}$$

where $\|\cdot\|_0$ denotes the l_0-norm of a vector, counting the number of non-zero entries in the vector. Donoho [41] proves that if matrix