Neurocomputing 205 (2016) 22–31

Contents lists available at ScienceDirect

Neurocomputing

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

Locality Constrained-ℓ_p Sparse Subspace Clustering for Image Clustering

Wenlong Cheng, Tommy W.S. Chow*, Mingbo Zhao

Department of Electronic Engineering, City University of Hong Kong, Hong Kong Special Administrative Region

A R T I C L E I N F O

Article history:
Received 29 May 2015
Received in revised form
10 March 2016
Accepted 20 April 2016
Communicated by Deng Cheng
Available online 17 May 2016

Keywords:
Subspace clustering
Sparse coding
ℓ_1-norm minimization
ℓ_p-norm minimization

A B S T R A C T

In most sparse coding based image restoration and image classification problems, using the non-convex ℓ_p-norm minimization $(0 \le p < 1)$ can often deliver better results than using the convex ℓ_1-norm minimization. Also, the high computational costs of ℓ_1-graph in Sparse Subspace Clustering prevent ℓ_1-graph from being used in large scale high-dimensional datasets. To address these problems, we in this paper propose an algorithm called Locality Constrained-ℓ_p Sparse Subspace Clustering (kNN-ℓ_p). The sparse graph constructed by locality constrained ℓ_p-norm minimization can remove most of the semantically unrelated links among data at lower computational cost. As a result, the discriminative performance is improved compared with the ℓ_1-graph. We also apply the k nearest neighbors to accelerate the sparse graph construction without losing its effectiveness. To demonstrate the improved performance of the proposed Locality Constrained-ℓ_p Sparse Subspace Clustering algorithm, comparative study was performed on benchmark problems of image clustering. Thoroughly experimental studies on real world datasets show that the Locality Constrained-ℓ_p Sparse Subspace Clustering algorithm can significantly outperform other state-of-the-art methods.

1. Introduction

In most machine learning and computer vision problems, data are often viewed as points lying in a union of multiple low dimensional subspaces, in which each subspace may correspond to one specific category or class, e.g., feature trajectories of moving objects captured by an affine camera [1], images of several subjects under varying illumination or under different poses [2], and local patches or texture features of pixels/superpixels of an image [3]. Subspace clustering [4], separates data points according to their underlying subspace, is one of the most widely used computational techniques for this kind of processing. For a given dataset obtained from a union of subspaces, subspace clustering finds the number of subspaces, determines the dimensionality of data, performs segmentation of the data and evaluates the basis for each subspace.

As performing fast online human face recognition [5], segmentation of human actions from videos [6] and handwriting types pattern recognition [7] problems become increasingly popular, various algorithms have been proposed to improve the performance of subspace clustering. Most of the early studies on subspace clustering are algebra or statistics based. Among algebra based methods, shape interaction matrix (SIM) [8] and generalized principal component analysis (GPCA) [9] are the two most well-known algorithms. However, the presence of noise, degeneracy, or partially coupled subspaces significantly affected the performance of this type of method. The statistics based methods, including random sample consensus (RANSAC) [10], expectation maximization (EM) [11], and several newly developed techniques like agglomerative lossy compression (ALC) [12], their performance are exquisitely dependent on the estimation of exact subspace models. Other different types of method using spectral clustering based [13] approaches have also been proposed. In particular, sparse representation [14] and low-rank approximation [15] based methods for subspace clustering have received considerable attention in recent years, nevertheless they do not require *a priori* knowledge of the dimensions and the number of subspaces. The segmentation of data is obtained by applying spectral clustering on the similarity graph based on sparse or low-rank representation. The Sparse Subspace Clustering (SSC) algorithm [14], which is well supported by theoretical analysis, provides state-of-the-art results on many widely used benchmark datasets. Elhamifar et al. [16] explored theoretical conditions to guarantee the correctness of clustering on noiseless data. Wang et al. [17] used a local neighborhood of each incomplete point to complete missing values, and refine the estimated subspaces to recover the full matrix. Soltanolkotabi et al. [18] presented a geometric function

* Corresponding author.
E-mail addresses: cheng.python@gmail.com (W. Cheng), eetchow@cityu.edu.hk (T.W.S. Chow), mzhao4@cityu.edu.hk (M. Zhao).

http://dx.doi.org/10.1016/j.neucom.2016.04.010
0925-2312/

analysis of SSC and proved that SSC succeeds when data corrupted by noise.

Sparse coding [19] represents a natural image as a sparse linear combination of atoms parsimoniously chosen out of an overcomplete dictionary. Results show that natural images can be generally coded by structural primitives (e.g., edges and line segments) that are qualitatively similar in a form to simple cell receptive fields [20]. Intuitively, using l_0-norm minimization (l_0-norm counts the number of nonzero entries in a vector) can measure the sparsity of the representation coefficient vector. But ℓ_0-norm minimization is an NP-hard problem which is challenging to solve because of the discontinuity and non-convexity of the ℓ_0-norm. Rather than solving the non-convex ℓ_0-norm minimization problem, one can replace ℓ_0-norm with its convex relaxation ℓ_1-norm $|x|_1 = \sum |x_i|$. The ℓ_1-norm minimization, a continuous and convex surrogate, has been proved is a good approximation for ℓ_0-norm minimization in finding the sparsest solution with high probability [21]. In fact, ℓ_1-norm minimization has been extensively studied [19] and applied to sparse representation applications. Although the ℓ_1-norm minimization based sparse learning formulations are able to deliver impressive results, recent experimental results show that ℓ_1-norm minimization is suboptimal [22], because the ℓ_1-norm, the closest convex approximation of the ℓ_0-norm, often leads to over-penalized problems. In order to address this issue, non-convex ℓ_p-norm $((0 \le p < 1)$ minimization, which interpolated between the ℓ_0-norm and the ℓ_1-norm, has been proposed for better approximation of the ℓ_0-norm [23]. Theoretical analysis and experimental results [24] suggest that the solution of ℓ_p-norm minimization is close to that of the ℓ_1-norm minimization and it is sparser. Recent theoretical studies have also demonstrated the superiority of ℓ_p-norm over the convex ℓ_1-norm in several sparse learning settings [25].

Other recent l_1-norm minimization work include Elhamifar and Vidal constructing a sparse similarity graph by using l_1-norm minimization based coefficients for subspace clustering, called Sparse Subspace Clustering (SSC) [14]. Without using a fixed global parameter to determine the size of neighborhood, SSC automatically reconstructs each datum from the remaining data by sparse coding. However, SSC solves the l_1-norm minimization for similarity graph construction but not the l_p-norm minimization. Moreover, for each new datum SSC has to perform the entire computational procedures over the whole dataset, which is very time-consuming and memory demanding for large-scale dataset. This makes SSC not suitable for fast online clustering.

In this paper, we propose an effective extension of SSC, called Locality Constrained-ℓ_p Sparse Subspace Clustering. Our proposed work first uses kNN to select the k nearest neighbors of each sample for the later sparse representation, and solves the l_p-norm minimization for similarity graph construction. We sparsely reconstruct each sample from its k nearest neighbors in feature space instead of using all the other samples to improve the efficiency while maintaining its effectiveness. And it is worth pointing out that our work is the first to directly employing l_p-norm sparse representation of vectors lying on a union of subspaces to cluster the data into separate subspaces. The l_p-norm minimization can yield solutions more sparser than those of l_1-norm minimization, and furthermore, can be efficiently solved by a simple iterative thresholding procedure. We then introduce an effective iterative shrinkage/thresholding method [26] to solve the l_p-norm minimization. By searching for the best sparse representation using kNN, our proposed algorithm can determine other points lying in the same subspace. This is important as it allows us to build a similarity matrix, from which segmentation of data can be subsequently obtained using spectral clustering.

Our Locality Constrained-ℓ_p Sparse Subspace Clustering method has the following advantages: (1) its l_p-norm minimization can remove most of the semantically-unrelated links to avoid the propagation of incorrect information than l_1-norm minimization, since each sample only has links to a small number of most probably semantically-related samples; (2) it is naturally more effective for discrimination since the kNN-ℓ_p-graph construction characterizing the local structure can convey important information for clustering; and (3) it is practical for large-scale applications because sparse representation can reduce the storage requirement while the approximate kNN-ℓ_p-graph construction is much more efficient than normal sparse graph construction. In this paper, we have conducted extensive experimental study on several real-world datasets. The presented results demonstrate the advantages of the proposed Locality Constrained-ℓ_p Sparse Subspace Clustering.

The rest of the paper is organized as follows: Section 2 provides a brief review of SSC and l_p-norm Minimization Technique. Section 3 presents the Locality Constrained-ℓ_p Sparse Subspace Clustering method. Section 4 carries out the experiments to examine the effectiveness of Locality Constrained-ℓ_p Sparse Subspace Clustering. Finally, Section 5 concludes this work.

2. Sparse Subspace Clustering and l_p-norm Minimization Technique

In this section, we start with a brief introduction of Sparse Subspace Clustering, ℓ_p-norm minimization sparse coding and then introduce a newly developed ℓ_p-norm minimization technique.

2.1. Notations

Assume we are given a collection of N data points where m indicates the input data dimension, denote the matrix containing all the data points as $Y = [y_1, \ldots, y_N] \in R^{m \times N}$, where each data point $y_i \in R^m$. Let $\{S_i\}_{i=1}^n$ be an arrangement of n subspaces, we assume data points $\{y_i\}$ lie in a union of n linear subspaces $S_1 \cup S_2 \cup \ldots S_n$. From the data points, we can construct a similarity graph represented as $G = (V, E, W)$. The vertices V denotes the data samples $\{y_1, y_2, \ldots, y_N\}$, and the edges E denotes the set of edges between nodes. The similarity matrix W with W_{ij} indicating the similarity between node y_i and node y_j, $N(y_i)$ represents a set of y_i's neighbors in graph G, excluding y_i and $K_i = |N_i|$. Given Y, the task of subspace clustering is to cluster data points according to their subspaces.

2.2. Sparse Subspace Clustering

Recently, researchers have utilized the inherent sparsity of sparse representation to construct a similarity graph for various tasks, e.g., dimension reduction [27], image analysis and other applications [28]. Among these works, Elhamifar and Vidal proposed their SSC algorithm for subspace clustering based on well-founded recovery theory for independent subspaces and disjoint subspaces. The motivation of SSC is that each data point x_i can be represented as a sparse linear combination of all the other data points within the same cluster, which is to learn a sparse coefficient matrix $X \in R^{N \times N}$. Formally, SSC solves the following convex l_1-norm minimization problem:

$$\min_X \ \frac{1}{2}\|Y - YX\|_F^2 + \lambda\|X\|_1 s.\ t.\ \mathrm{diag}(X) = 0. \quad (1)$$

where diag(X) is the diagonal vector of matrix X.

The sparse matrix X has two important usages. First, for each datum y_i, its neighbors N_i can be easily inferred by the nonzero elements in the i-th column of X. We can theoretically guarantee