Knowledge-Based Systems 121 (2017) 129–141

Contents lists available at ScienceDirect

Knowledge-Based Systems

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

Efficient locality weighted sparse representation for graph-based learning

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

[a] School of Political Science and Public Administration, University of Electronic Science and Technology of China, Chengdu, Sichuan 611731, China
[b] Donlinks School of Economics and Management, University of Science and Technology Beijing, Beijing 100083, China
[c] Business Analytics and Statistics, University of Tennessee, Knoxville, TN 37996, USA

ARTICLE INFO

Article history:
Received 6 July 2016
Revised 10 December 2016
Accepted 14 January 2017
Available online 18 January 2017

Keywords:
Sparse representation
Locality preserving
Graph-based learning
Graph construction

ABSTRACT

Constructing a graph to represent the structure among data objects plays a fundamental role in various data mining tasks with graph-based learning. Since traditional pairwise distance-based graph construction is sensitive to noise and outliers, sparse representation based graphs (e.g., ℓ_1-graphs) have been proposed in the literature. Although ℓ_1-graphs prove powerful and robust for many graph-based learning tasks, it suffers from weak locality and high computation costs. In this paper, we propose a locality weighted sparse representation (LWSR), which aims for good preservation of the locality structure among data objects and a significant reduction of the computation time. LWSR approximates each object as a sparse linear combination of its nearest neighbors, and weights their corresponding coefficients by their distances to the target object. Experimental results show that LWSR-graph based learning methods outperform state-of-the-art methods in both effectiveness and efficiency for graph-based learning.

1. Introduction

Graph-based learning is a popular class of approaches that can be used in many data mining tasks, such as unsupervised clustering [1], semi-supervised learning [2], and manifold learning [3]. For graph-based learning, a critical step is to construct a directed or undirected graph that represents the proximity structure among data objects. In traditional graphs, such as ε-ball neighbors, k-nearest neighbors, and fully-connected Gaussian Radial Basis Function (RBF) graphs [4], the edges and weights among data objects are determined by pairwise Euclidean distances in the original data space. However, such distances are sensitive to noise and outliers, and are subject to the curse of dimensionality [5]. In other words, data with noise, outlying objects, and high-dimensions tend to dramatically distort the structure of the graph.

A widely used principle for noise reduction in high-dimensional space is to construct sparse estimates. The ℓ_1-graph [6], for example, is based on a modified sparse representation framework, which represents each data object by a sparse linear combination of all other data objects [7]. A series of algorithms based upon the ℓ_1-graph are obtained by using ℓ_1-regularization or ℓ_1-minimization, such as sparse subspace clustering [1], label propagation [8], sparse reconstruction embedding [9], sparse representation based linear projections [10], and sparsity preserving projections [11].

Although it has been shown that these sparse representation based graphs outperform pairwise distance-based graphs in most graph-based learning tasks or approaches [6], they do have limitations. One limitation is that sparse representation using ℓ_1-regularization for constructing graphs only guarantees the sparsity property, and does not preserve the locality. In other words, the locality structure in the original space is lost in the coefficient space. Another limitation is that a non-smooth convex problem must be solved whose computational complexity is proportional to the cube of the problem size. When the size of a dataset is large, the computation cost of solving the sparse representation becomes prohibitively high.

In this paper, we propose a locality weighted sparse representation (LWSR), which is expected to preserve the locality structure and consume much less computation time than the original sparse representation. More specifically, LWSR selects a set of local objects and obtains the sparse representation over these selected local objects only. Unlike ℓ_1-graph, which optimizes the sparse coefficients over all other objects, LWSR's local embedding with respect to nearest neighbors has the advantage of consuming much less computation time while capturing most of the information. Furthermore, LWSR weights the reconstruction coefficients by the neighbored objects' distances to the target object. By

* Corresponding author.
E-mail address: fengxd1988@hotmail.com (X. Feng).

http://dx.doi.org/10.1016/j.knosys.2017.01.019
0950-7051/

integrating the LWSR-graph into popular graph-based learning frameworks, including spectral clustering, subspace learning, and semi-supervised learning, we demonstrated the effectiveness of the proposed method with comprehensive experiments. Our contribution can be highlighted as follows.

- We proposed a novel scheme (LWSR) for finding sparse similarity graphs by reconstructing the target object with its neighbors under sparse representation framework.
- The proposed LWSR can well balance locality and sparsity in the graph; and it is to reduce the computation complexity of $\ell_1 - graph$ much.
- We show theoretical analysis to prove that the LWSR-graph must be block diagonal under some conditions,and thus it can well reveal the true subspace membership.
- The proposed algorithm is robust to various corruptions and real occlusions in the context of subspace clustering, subspace learning, and semi-supervised learning.
- Theoretical analysis and experiments demonstrate that the main advantage of the proposed sparse representation based graph is the much lower computational complexity and for subspace learning it also archives higher accuracy than all counterparts due to the well locality preservation.

The rest of the paper is organized as follows. Section 2 presents some preliminaries and related work on graph construction methods. In Section 3, the proposed LWSR method is described and discussed in detail. We derive a series of methods using LWSR based graphs in Section 4, including spectral clustering, subspace learning, and semi-supervised learning. Section 5 summarizes experimental results to evaluate the effectiveness and efficiency of the proposed algorithm in graph-based learning. Finally, we draw conclusions and point out possible future work in Section 6.

2. Preliminaries and related works

In this section, we review and discuss existing graph-based learning methods using sparse representation. It is noteworthy that although the intentions of different graph-based learning algorithms vary, they have a similar purpose to construct a graph structure that models the contextual information of the data manifold. Suppose we have a dataset $X = [\vec{x}_1, \vec{x}_2, \ldots, \vec{x}_n] \in \mathbb{R}^{m\times n}$ consisting of n data objects, where $\vec{x}_i = [x_{i1}, x_{i2}, \ldots, x_{im}]^T \in \mathbb{R}^m$ is a vector representing the ith data object. With these data objects, a graph $G = (V, E)$ may be built, where the vertex set V is the set of all data objects in X, and the edges in E may be weighted or unweighted, depending on the graph construction method.

2.1. Sparse representation

In order to simultaneously determine the graph adjacency structure and the graph weights, sparse representation is exploited to obtain a sparse coefficient vector for each data object. Coming from compressed sensing, sparse representation makes the assumption that each object can always be approximately represented by a linear combination of other objects, where the coefficients should be sparse. Sparse representation has a wide application in image analysis [12,13]. Motivated by the fact that an ideal affinity matrix is block diagonal and sparse, sparse representation based graphs, such as sparse subspace clustering [1], sparsity induced similarity (SIS) [8], ℓ_1 directed graph [2], and ℓ_1-graph [6], attempt to solve the following sparse representation problem as an optimization problem for each data object:

$$\min_{\vec{\alpha}_i} \|\vec{\alpha}_i\|_0, \ \text{s.t.}\ \vec{x}_i = X_{-i}\vec{\alpha}_i, \tag{1}$$

where $\|\cdot\|_0$ denotes the ℓ_0-norm of a vector, counting the number of non-zero elements in the vector; X_{-i} is the data matrix that omits $\vec{x}_i$; and $\vec{\alpha}_i = [\alpha_{i,1}, \ldots, \alpha_{i,i-1}, \alpha_{i,i+1}, \ldots, \alpha_{i,n}]^T \in \mathbb{R}^{n-1}$ is the sparse representation of $\vec{x}_i$ over all other data objects. Donoho proved that if a matrix satisfies the restricted isometry property, Eq. (1) has a unique solution [14]. However, solving Eq. (1) is NP-hard, as the objective function is non-convex. In other words, there is no known approach to find the sparsest solution that is significantly more efficient than exhausting all subsets of the components for $\vec{\alpha}_i$. Researchers in emerging theories of compressed sensing reveal that the non-convex optimization in Eq. (1) is equal to the following convex ℓ_1 optimization problem, if the optimal $\vec{\alpha}_i$ is sparse enough [15]:

$$\min_{\vec{\alpha}_i} \|\vec{\alpha}_i\|_1, \ \text{s.t.}\ \vec{x}_i = X_{-i}\vec{\alpha}_i, \tag{2}$$

where $\|\cdot\|_1$ denotes the ℓ_1-norm of a vector, summing the absolute value of each entry in the vector. This problem can be solved in polynomial time using convex programming tools, and is known to prefer sparse solutions [16].

The solution of sparse representation can also be written for all objects in matrix form as:

$$\min_{A} \|A\|_1, \ \text{s.t.}\ X = XA \ \text{and}\ \operatorname{diag}(A) = \vec{\mathbf{0}}, \tag{3}$$

where $A \in \mathbb{R}^{n\times n}$ is the sparse representation matrix of the dataset over itself; $\|A\|_1$ is the ℓ_1-norm of the matrix A, summing the absolute value of each element; the ith column vector $\vec{\alpha}_i$ in A denotes the sparse representation of $\vec{x}_i$; $\operatorname{diag}(\cdot)$ is the vector of the diagonal elements of a square matrix; and $\vec{\mathbf{0}} \in \mathbb{R}^n$ denotes a vector with all zero elements.

Since real world data contain noise, it may not be possible to express each data object exactly as a sparse representation of the other data objects. Therefore, the sparse solution A can be approximately obtained by solving the following ℓ_1-optimization problem:

$$\min_{A} \|A\|_1 + \lambda\|Z\|_F, \ \text{s.t.}\ X = XA + Z \ \text{and}\ \operatorname{diag}(A) = \vec{\mathbf{0}}, \tag{4}$$

where $\|\cdot\|_F$ represents the Frobenius norm of a matrix [1]; $Z \in \mathbb{R}^{m\times n}$ corresponds to the matrix of reconstruction error; and $\lambda > 0$ denotes the regularization parameter. It is worth noting that Eq. (4) is widely known as the least absolute shrinkage and selection operator optimization (LASSO) [17].

2.2. Graph construction

When it comes to graph construction, an interpretation for the sparse coefficient α_{ij} is how much data object $\vec{x}_j$ contributes to the reconstruction of $\vec{x}_i$. The vector $\vec{\alpha}_i$ can be treated as a vector of contribution weights from all data objects to the reconstruction of $\vec{x}_i$. Therefore, the weight matrices of the proximity graph may be defined in various ways, such as:

- $W = A$ as in ℓ_1-graph [6];
- $W = |A|$ as in ℓ_1 directed graph [2];
- $W = \frac{\tilde{A}+\tilde{A}^T}{2}$ as in SIS [8], where $[\tilde{A}]_{ij} = \frac{\max\{0,\alpha_{ij}\}}{\sum_k \max\{0,\alpha_{ik}\}}$; and
- $W = \max\left\{0, \frac{\vec{\alpha}_i\cdot\vec{\alpha}_j}{\|\vec{\alpha}_i\|_2\times\|\vec{\alpha}_j\|_2}\right\}$ as in [18].

Based on Eqs. (1)–(4), it is clear that all these sparse representation based graphs aim to obtain a similarity graph by solving an ℓ_1-minimization problem. In spite of the success of the sparse representation based graphs in several important applications [19,20], the locality structure is not well preserved. More specifically, the goal of sparse representation is to learn sparse weights over other data objects. Weights may be allocated to a distant data object to favor sparsity.

Besides the problem of loss of locality in sparse representation, another drawback is the computation cost for optimizing the reconstruction coefficients. It is known that the objective function of