Neurocomputing 000 (2018) 1–10

Contents lists available at ScienceDirect

Neurocomputing

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

Symmetric low-rank representation with adaptive distance penalty for semi-supervised learning

Chang-Peng Wang[a,*], Jiang-She Zhang[b], Fang Du[b], Guang Shi[b]

[a] School of Science, Chang'an University, Xi'an 710064, China
[b] School of Mathematics and Statistics, Xi'an Jiaotong University, Xi'an 710049, China

ARTICLE INFO

Article history:
Received 12 December 2017
Revised 10 May 2018
Accepted 8 August 2018
Available online xxx

Communicated by Dr Zhao Zhang

Keywords:
Low-rank representation
Semi-supervised learning
Graph construction
Image classification

ABSTRACT

Graph-based approaches have been successfully used in semi-supervised learning (SSL) by weighting the affinity between the corresponding pairs of samples. Low-rank representation (LRR) is one of the state-of-the-art methods, has been extensively employed in the existing graph-based learning models. In graph-based SSL, a weighted affinity graph induced by LRR coefficients is used to reflect the relationships among data samples. Constructing a robust and discriminative graph to discover the intrinsic structures of the data is critical for SSL, because the central idea behind the graph-based SSL approaches is to explore the pairwise affinity between data samples to infer those unknown labels. However, most of existing LRR-based approaches fail to guarantee weight consistency for each pair of data samples, which is beneficial to preserve the subspace structures of high dimensional data. In this paper, we propose a symmetric LRR with adaptive distance penalty (SLRRADP) method for the small sample size (SSS) recognition problem. The graph identified by SLRRADP can not only preserve global and local structures of data, but also exploit intrinsic correlation information of data samples. Extensive experimental results on semi-supervised classification tasks demonstrate the effectiveness of the proposed method and its superiority in performance over the state-of-the-art graph construction approaches.

1. Introduction

In real-world applications of machine learning and pattern recognition, it is often the case that abundant unlabeled samples are available, but the insufficient labeled samples are hard or expensive to acquire. As a consequence, semi-supervised learning (SSL) [1–3], which attempts to simultaneously utilize both limited labeled samples and rich unlabeled ones, has attracted increasing attention in learning-based applications. The target goal of SSL algorithms is to build the connection between unlabeled data and labeled data, and then predict the labels of the unlabeled data by semi-supervised classifier. During the past decade, a variety of SSL algorithms have been developed, e.g., self-training [4], co-training [5], boosting-based SSL [6], graph-based SSL [7] and deep learning-based SSL [8].

Graph is an important structure for describing data relationship in many learning tasks. Based on the manifold assumptions [1,9], that is, nearby points are likely to have the same label, recent efforts have been made on the graph-based SSL (GSSL) [10–13], which attempt to better exploit the intrinsic (either local or global) geometrical structure inferred from both labeled and unlabeled data. With a faithful graph, GSSL obtains many advantages such as closed-form solution, promising learning performance, and thus GSSL attracts significant attention since it was proposed and has been widely applied in many real-world applications [14–16].

However, it is widely known that the graph quality, rather than the learning/optimization algorithm, seriously affects the performance of GSSL methods. Some techniques based on Frobenius norm are developed to construct ℓ_2-graph with the advantage of low computational cost [17,18]. Besides, Elhamifar and Vidal [19] proposed a sparse subspace clustering (SSC) method, which utilizes the sparsest representation of the data via ℓ_1-norm minimization of the coefficient matrix to define an affinity matrix. Liu et al. [20] proposed a large scale distributed framework for the computation of SSC via an alternating direction method of multiplier (ADMM) algorithm. However, there is no global structure constraint on the sparse representation in SSC.

Unlike SSC which enforces the representation coefficients to be sparse, the recently proposed LRR can obtain a low-rank coefficient Z by solving a rank minimization problem [21]. LRR can capture the global structure of data and perform robust performance for the corrupted data. The graph (affinity matrix) constructed by Z has been widely used for many learning tasks, such

* Corresponding author.
E-mail address: cpwang@chd.edu.cn (C.-P. Wang).

https://doi.org/10.1016/j.neucom.2018.08.017
0925-2312/

as subspace segmentation [22,23], image clustering [24,25] and classification [26,27]. Several improved models have been proposed in order to construct a robust and discriminative graph for SSL. Latent low-rank representation (LatLRR) [23] is proposed to robustly extract salient features from corrupted data. Robust LatLRR [28] first denoises the data by robust PCA and then finds the low-rank and sparse solution. LRR with adaptive distance penalty (LRRADP) [29] can not only capture the global subspace structure of the whole data, but also consider the neighbor relationship among samples.

LRR-based approaches have achieved excellent performance in respective applications, and have been proved to be effective at handling corrupted data. However, they usually define an affinity matrix by a symmetrization step $W = (|Z| + |Z^T|)/2$. When the entries of the matrix Z are sensitive to the noise or outliers in low-rank representation, it inevitably leads to lose intrinsic correlation information of samples. Thus, the graph constructed by $W = (|Z| + |Z^T|)/2$ is not able to accurately characterize how other samples contribute to the reconstruction of a given sample. To overcome this drawback, we attempt to integrate the symmetric constraint into the framework to learn a more robust and effective LRR model. In this paper, we propose a symmetric low-rank representation with adaptive distance penalty (SLRRADP) method, which is not only preserve the global subspace structures and the intrinsic neighbor relationships of high-dimensional data, but also exploit the geometrical structure of the memberships of data samples. In SLRRADP, the global structure is considered by the low-rank property and the local structure is emphasized by the adaptive distance penalty. Besides, the symmetric constraint could guarantee the weight consistency for each pair of data samples, and thus preserve the multiple subspace structure. With the affinity matrix built by the SLRRADP, the conventional semi-supervised classification method Gaussian Fields and Harmonic Functions (GFHF) [30], can be used for predicting the label of unlabeled samples. In contrast to the conventional LRR-based approaches, SLRRADP can obtain a more accurate coefficient matrix, and then SLRRADP could build a better affinity matrix for image classification. Further details will be discussed in Section 3.

In summary, our main contributions in this paper lie in the following three aspects:

(1) Unlike conventional LRR-based methods that obtaining an optimal solution and defining a symmetric affinity matrix are treated separately, SLRRADP integrates these two tasks into a single optimization framework to guarantee that the intrinsic correlation information of data samples can be preserved.
(2) The low-dimensional representation obtained by SLRRADP exploits the global subspace structure (based on the LRR), the local structure (by the adaptive distance penalty), and the structure of the memberships of data samples (by the symmetric constraint), which could be used to learn a robust and discriminative affinity matrix.
(3) Extensive experimental results on benchmark databases demonstrate the superiority of the proposed method for the small sample size (SSS) recognition problem.

The rest of the paper is organized as follows. Section 2 mainly reviews the related works. Section 3 introduces and analyzes the details of our proposed method SLRRADP. Experiments are conducted and presented in Section 4. Finally, we provide the conclusions in Section 5.

2. Related works

The goal of SSL is to exploit the structure of the data by using both labeled and unlabeled samples. Building a good connection to measure the similarity between unlabeled data and labeled data is critical in SSL, because it attempts to identify the labels of the unlabeled data according to the affinity between the pairs of samples. The recently LRR is a promising weight graph construction method. This section provides a review of low-rank representation model used in semi-supervised classification framework, which is closely related to the proposed method.

Before reviewing previous work, we define some notations. Let $X = [X_1, X_2, \ldots, X_n] \in \mathbb{R}^{d\times n}$ be a set of n samples in the d-dimensional space. For a matrix X, X_i is its ith column and X_{ij} is its (i, j)th entry. The trace of X is denoted as $tr(X)$, and X^T is the transpose of X. The ℓ_p norm of matrix X is denoted as $\|X\|_p$. Specially, $\sigma_i(X)$ is defined as the ith singular value of X, $\|X\|_* = \sum_i \sigma_i(X)$ and $\|X\|_1 = \sum_{i,j} |X_{ij}|$ denote the nuclear norm and ℓ_1 norm, respectively. $\langle A, B\rangle = tr(A^T B)$ represents the inner product, which is equal to the sum of the diagonal elements of the matrix $A^T B$.

2.1. Classification by low-rank representation

Recently, some theoretical advances on LRR enable us to explore low-dimensional subspace structures embedded in data. Given a set of data samples, LRR aims at finding the lowest-rank representation of all samples jointly and preserving the membership of samples that belong to the same subspace [21]. Thus, the sample usually can be represented by other samples that lie in the same subspace when the subspace are independent and the samples are noiseless. Generally, LRR solves the following nuclear norm minimization problem for the noise free case

$$\min_Z \|Z\|_* \quad \text{subject to} \quad X = AZ \tag{1}$$

where the columns of A are a set of known bases or dictionary items and Z is called the low-rank representation of data X. $\|\cdot\|_*$ denotes the nuclear norm (the sum of its singular values of the matrix), which is the convex envelope of the the rank function [31].

In the real world applications, the observation data often contain noise corruption, and the data matrix X itself is usually used as the dictionary in LRR to automatically correct corruptions during data recovery. With the balance parameter λ, a more reasonable objective function for LRR can be expressed as

$$\min_{Z,E} \|Z\|_* + \lambda\|E\|_l \quad \text{subject to} \quad X = XZ + E \tag{2}$$

Here, there are many strategies to define the error term E. For example, ℓ_0-norm characterizes the random corruption, $\ell_{2,1}$-norm generally character sample-specific corruption, and F-norm is proposed for the small Gaussian noise. A number of methods have been proposed for solving the above low-rank matrix problems, and the most common used methods are augmented Lagrange multiplier (ALM) method and its variants [32,33]. The advantage of LRR mainly comes from the low-rank component could reduce the influence of the outliers, which means LRR has the ability to correct the corruptions in data automatically.

After obtaining an optimal solution (Z^*, E^*), Z^* is used to define an affinity matrix $W = (|Z^*| + |Z^*|^T)/2$ to measure the similarity between the data samples. Then, the semi-supervised classifier, such as Local and Global Consistency (LGC) [9] or Gaussian Fields and Harmonic Functions (GFHF) [30], can be used to predict labels of unlabeled samples based on the affinity matrix.

Gaussian Fields and Harmonic Functions (GFHF) [30] is a popular semi-supervised classifier, and will be used to predict labels of unlabeled samples in this paper. GFHF utilizes the weight graph W and labeled matrix Y to recover the continuous classification function by simultaneously optimizing both the label fitness and manifold smoothness. The optimization function of GFHF is to recover

Please cite this article as: C.-P. Wang et al., Symmetric low-rank representation with adaptive distance penalty for semi-supervised learning, Neurocomputing (2018), https://doi.org/10.1016/j.neucom.2018.08.017