Neurocomputing 140 (2014) 84–96

Contents lists available at ScienceDirect

Neurocomputing

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

A simplified low rank and sparse graph for semi-supervised learning ☆

Miaoyun Zhao *, Licheng Jiao, Jie Feng, Tianyu Liu

Key Laboratory of Intelligent Perception and Image Understanding of Ministry of Education, International Research Center for Intelligent Perception and Computation, Xidian University, Xi'an, Shaanxi Province 710071, China

ARTICLE INFO

Article history:
Received 11 September 2013
Received in revised form
14 March 2014
Accepted 15 March 2014
Communicated by Shiliang Sun
Available online 18 April 2014

Keywords:
Low rank representation
Sparse representation
Sparse congruency representation
Joint sparse representation

ABSTRACT

Low rank representation is capable of capturing the global structure of mixed subspaces which are usually assumed to be independent. However, its computation is time-consuming. In practice, the data always distributes on subspaces that intersect or even overlap with each other. So the local structure of the data among the overlapping parts is important. Sparsity is a good property to accelerate the algorithm and capture the local linear structure. The drawback is that it breakups the low rank property of the reconstruction coefficient matrix when combining with low rank representation. In order to combine the two advantages properly, in this paper, we introduce a new constraint to the low rank representation matrix, which is called sparse congruency. Fortunately, we find that this new constraint is a simplification to the low rank and sparse constraints. Several experiments are implemented to demonstrate the efficiency of our method in semi-supervised classification.

1. Introduction

With rapid development of the Internet, we are drowned in the massive unlabeled data. How to make thorough use of the mass of data becomes more and more important. In the application of machine learning, such as pattern recognition, object detection and image searching, we have to face the problem of lacking sufficient labeled training data because the work of labeling is really costly and laborious. Semi-supervised learning (SSL) [1] is a technique which tries to cope with the very situations of limited labeled data and abundant unlabeled data.

Recently, SSL has received a significant amount of attention in both theory and practice. A great number of SSL methods, such as Expectation-Maximization (EM) with generative mixture models, self-training, co-training, transductive Support Vector Machine (TSVM) [1–5], Sparse SSL using conjugate functions [6], Multi-view Laplacian Support Vector Machines (SVMs) [7], and graph-based methods, have been widely used and referenced. We will pay more attention to the graph-based SSL method [8–11] in this paper since its empirical success in practice and the computational efficiency. Graph-based SSL methods define a graph $G=(V,E)$, where nodes V are the labeled and unlabeled samples in the dataset, edges E are associated with a weight matrix W which reflects the similarity among samples. The label information then can be efficiently and effectively propagated to unlabeled samples across the graph. These methods, which are nonparametric, discriminative, and transductive in nature, usually assume label smoothness over the graph. According to literatures [12,13], the construction of the graph should try to accommodate the cluster assumption [3]: points on the same structure are likely to share the same label. Therefore, the critical thing for graph-based SSL is to correctly construct a graph which can best capture the essential data structures [14,15]. Low rank representation (LRR) is testified to significantly outperform the other state of the art methods at exactly recovering the subspace structure of the original dataset, and robust to noise [16]. A given dataset can be reasonably described as points lying on several separated or mixed subspaces. The problem to cluster data into groups with each group corresponding to a subspace is usually called the subspace segmentation problem [17]. Comparing to the other subspace methods (i.e. Generalized Principal Component Analysis (GPCA) [18], Robust Algebra Segmentation (RAS) [19], Normalized Cuts (NCut) [20], Sparse Subspace Clustering (SSC) [21], and Spectral Curvature Clustering (SCC) [22]), LRR is better at capturing the global structures of the dataset. So LRR is more suitable to construct a good graph for SSL in the subspace segmentation problem.

An informative graph should have the following three characteristics [23]: high discriminating power, high sparseness, and adaptive neighborhood. Though LRR is excellent to construct a good graph for SSL in subspace segmentation problem, its computation is time-consuming and it always results in a comparably dense graph which is time-consuming to analyze. A common problem of graph based methods in the area of machine learning is that the complexity of obtaining the solution scales as $O(N^3)$, in

☆This document is a collaborative effort.

* Corresponding author. Tel.: +86 15109284921.
E-mail addresses: zhao_miaoyun@163.com (M. Zhao), lchjiao@mail.xidian.edu.cn (L. Jiao), jiefeng0109@163.com (J. Feng), tianyu_liu@163.com (T. Liu).

http://dx.doi.org/10.1016/j.neucom.2014.03.033
0925-2312/

which N is the data size. To accelerate the computation, literature [24] proposed a typical low rank approximation technique, named Nystrom method, for Large-scale machine learning problems. With a small part of columns sampled from the original Symmetric Positive Semi-Definite (SPSD) matrix, Nystro m method approximates the original SPSD matrix by the following SPSD Sketching Model [24]: Let A be an $n \times n$ positive semi-definite matrix, and S be a matrix of size $n \times l$, where $l \ll n$. Take $C=AS$ and $W = S'AS$. Then $CW^{\dagger}C'$ is a low-rank approximation to A with rank at most l. Nystrom method is usually used to approximate the affinity matrix in the area of machine learning. The performance of Nystrom method is greatly affected by the way of sampling columns from the original matrix (namely, the structure of matrix S). Yan et al. [25,26] proposed a sparse graph based on sparse representation (SR), which is sparse datum adaptive and robust to noise. Recently, literature [27] also proposed a sparse method, named manifold-preserving graph reduction, for sparse SSL. The only shortage of this method is that the sparse graph method finds the sparsest representation for each sample individually and thus it lacks a global consistent constraint on the solutions. So the representation bases of all samples will be so multifarious that we cannot see the subspace structure of the dataset. SR is a global method to one sample, but the global structure to be captured here is the structure reflected by the representations of all the samples in the dataset. Fortunately, Zhuang et al. [28] proposed a Non-Negative Low Rank and Sparse Graph (NLS1-graph) for SSL. By combining the sparse and low rank constraints on the reconstruction coefficient matrix, the NLS1-graph informatively harnesses both sparse and low rank properties of the dataset. Thus, the NLS1-graph can capture not only the local low-dim linear relationship but also the global subspace structure of the dataset. However, directly combining of sparse and low rank constraints seems to be too arbitrary. The sparse constraint for the reconstruction coefficient matrix is entry-wise, resulting in the lacking of a global constraint. Thus, the sparse constraint will break up the low rank property of the reconstruction coefficient matrix, and more iterations are needed for the algorithm to get convergence. Another small shortage is that NLS1-graph needs two independent parameters to trade off among three constraints: low rank, sparse, and noise (reconstruction error). Thus, the manual choose of the parameters is really difficult.

In this paper, we did two contributions.

First, we introduce a constraint from the literature [29,30], which is called joint sparse representation, to settle the problem existing in the combination of LRR and SR. In contrast with the NLS1-graph method, we cut the small valued entries of a reconstruction coefficient matrix without breaking up the low rank property. The proposed method is suitable for both the cases that samples of different classes exist in different subspaces and the cases that samples exist in two overlapped subspaces or independent subspaces. We will see in the experiments part that when there are some samples exist in two overlapped subspaces, though the performance of our method is affected, it enhanced the performance of LRR based methods on overcoming the subspace overlapping problem and kept the best one comparing to the LRR based methods. There is a lot work to do to more perfectly overcome this shortage. For simplicity, we assume that the reconstruction coefficient vectors for the samples of the same cluster share a common sparse pattern. And at the same time, the values of the coefficients corresponding to the same atom may be different for each sample. That is to say, the joint sparse representation allows all the samples belonging to the same cluster to be reconstructed by the same small set of atoms, while weighted with specific values. Thus, the reconstruction coefficient matrix is possible to get sparse without losing its low rank property.

Second, we observed that the joint sparse representation has two properties: low rank and sparse. That is to say, the joint sparse representation itself is sufficient to capture both the local and global structure information of the dataset. Here, for convenience, we call the constraint used in the joint sparse representation as sparse congruency constraint. And the representation method under this constraint is named as sparse congruency representation (SCR). Motivated by this idea, we proposed a novel method to construct an informative graph, which is low rank and sparse, with only one parameter to balance between two constraints: sparse congruency and noise (reconstruction error). We call such graph as sparse congruency graph.

Extensive experiments have been conducted on public databases for the semi-supervised classification problem. In the experiments part, we will see that the sparse congruency graph has significantly improved the performance of SSL, especially when it is compared with the classical low rank based methods. All these results remarkably demonstrate that our method can construct a more informative graph with less time than the classical methods.

Our work has several conceptual advantages as follows:

First, the sparse congruency constraint ensures the graph to be sparse and low rank at the same time. Thus the graph captures the local and global structure of the dataset simultaneously without facing the confliction between SR and LRR;

Second, when the subspaces are overlapped with each other, the low rankness of the dataset seems inconspicuous. So the local structure becomes important and accurate in exploring the cluster structure of the dataset, while the low rank constraint may mislead us. This is because LRR is more suitable for the cases that samples of different classes exist in different subspaces. When the subspaces are overlapped, it is difficult for LRR to identify the subspaces of different clusters. Thus the sample in a specific subspace can be represented by the samples within this subspace as well as subspaces which are overlapped with this subspace. For example, Fig. 2(d) shows a reconstruction coefficients matrix obtained by LRR. The dataset belongs to three clusters. The subspace of the third cluster is overlapped with the subspaces of the first two clusters. We can observe that samples in the third cluster have a close relationship with the samples in all the three clusters. Compared with the classical low rank based methods, the sparse congruency constraint integrates the low rank property into SR skillfully and avoids the misleading of LRR. Thus a proper local low-dimensional linear relationship of the data can be exploited.

Finally, due to the abandoning of the low rank constraint, our method saves a great amount of time computing the singular value decomposition (SVD). And we have only one parameter to tune mutually. What is more, the skillful resolution of the confliction between sparse and low rank properties saves a lot of time, too.

The rest of the paper is organized as follows. In Section 2, we precisely illustrate the problem trying to be solved. In Section 3, the key idea of our method is explained and the details on the formulation of a SCR optimization problem are presented. In Section 4, an algebraic solution for the optimization problem is displayed and summarized. In Section 5, we give the detailed procedure on how to construct the sparse congruency graph and how to integrate the graph into the SSL framework. In Section 6, extensive experiments and detailed analysis are presented. Finally, we conclude our paper in Section 7.

2. Background

LRR has many advantages over the other manifold learning methods when used to capture the global structure of the dataset. However, its drawbacks are not neglectable: First, LRR is not able to maintain the local feature which is as important as the global feature when the subspaces overlapped with each other; second, when the dataset is too big or the feature dimension is too large,