Neural Networks 65 (2015) 1–17

Contents lists available at ScienceDirect

Neural Networks

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

Enhanced low-rank representation via sparse manifold adaption for semi-supervised learning

Yong Peng [a], Bao-Liang Lu [a,b,*], Suhang Wang [c]

[a] Center for Brain-like Computing and Machine Intelligence, Department of Computer Science and Engineering, Shanghai Jiao Tong University, Shanghai 200240, China

[b] Key Laboratory of Shanghai Education Commission for Intelligent Interaction and Cognitive Engineering, Shanghai Jiao Tong University, Shanghai 200240, China

[c] Department of Computer Science and Engineering, Arizona State University, Tempe, AZ 85281, USA

ARTICLE INFO

Article history:
Received 3 December 2013
Received in revised form 2 December 2014
Accepted 4 January 2015
Available online 10 January 2015

Keywords:
Low-rank representation
Sparse manifold adaption
Graph construction
Semi-supervised learning
Face recognition

ABSTRACT

Constructing an informative and discriminative graph plays an important role in various pattern recognition tasks such as clustering and classification. Among the existing graph-based learning models, low-rank representation (LRR) is a very competitive one, which has been extensively employed in spectral clustering and semi-supervised learning (SSL). In SSL, the graph is composed of both labeled and unlabeled samples, where the edge weights are calculated based on the LRR coefficients. However, most of existing LRR related approaches fail to consider the geometrical structure of data, which has been shown beneficial for discriminative tasks. In this paper, we propose an enhanced LRR via sparse manifold adaption, termed manifold low-rank representation (MLRR), to learn low-rank data representation. MLRR can explicitly take the data local manifold structure into consideration, which can be identified by the geometric sparsity idea; specifically, the local tangent space of each data point was sought by solving a sparse representation objective. Therefore, the graph to depict the relationship of data points can be built once the manifold information is obtained. We incorporate a regularizer into LRR to make the learned coefficients preserve the geometric constraints revealed in the data space. As a result, MLRR combines both the global information emphasized by low-rank property and the local information emphasized by the identified manifold structure. Extensive experimental results on semi-supervised classification tasks demonstrate that MLRR is an excellent method in comparison with several state-of-the-art graph construction approaches.

1. Introduction

For many machine learning and pattern recognition applications, we often have no sufficient labeled samples, which are usually hard or expensive to acquire. However, unlabeled samples are easier to obtain via the Internet for some applications. For simultaneously utilizing both limited labeled samples and many unlabeled samples, SSL has received increasing attention in learning-based applications. SSL algorithms usually make use of the smoothness, cluster, and manifold assumptions, which can be roughly categorized into four groups: generative models, low-density separation models, heuristic models, and graph-based models. Nie et al. presented a semi-supervised orthogonal discriminant analysis algorithm via label propagation by solving the orthogonal constrained trace ratio optimization problem (Nie, Xiang, Jia, & Zhang, 2009). Yu et al. proposed a two stage method in which an unsupervised basis learning phase was followed by a supervised function learning, for SSL on high dimensional nonlinear manifolds (Yu, Zhang, & Gong, 2009). A unified framework for semi-supervised and unsupervised dimensionality reduction was proposed in Nie, Xu, Tsang, and Zhang (2010). A SSL framework, termed flexible manifold embedding, considers the manifold structure of both labeled and unlabeled samples. Karasuyama et al. designed a parameterized similarity function to define the graph edge weights (Karasuyama & Mamitsuka, 2013), which represent both similarity and local representation weight simultaneously. A detailed review of recent work on SSL can be found in Zhu (2008). In this paper, we focus our work on graph-based SSL due to its excellent performance in practice.

* Corresponding author at: Center for Brain-like Computing and Machine Intelligence, Department of Computer Science and Engineering, Shanghai Jiao Tong University, Shanghai 200240, China.
E-mail addresses: stany.peng@gmail.com (Y. Peng), bllu@sjtu.edu.cn (B.-L. Lu).

http://dx.doi.org/10.1016/j.neunet.2015.01.001
0893-6080/

Graph-based SSL relies on using a graph $G = (V, E, W)$ to represent the data structure, where V is a set of vertices and each vertex represents a data point, $E \subseteq V \times V$ is a set of edges connecting related vertices, and W is an adjacency matrix measuring the pairwise weights between vertices. Generally, the graph is constructed by using the relationship of domain knowledge or similarity of samples. Once the graph is constructed, each sample spreads its label information to neighbors over the graph until a global stable state is achieved on the whole data set. Thus, both labeled and unlabeled samples remarkably affect the graph construction. How to construct a good graph for representing data structure is critical for graph-based SSL. Recently, some graphs have been well investigated, such as k nearest neighbors (KNN) graph, local linearly embedding (LLE)-based graph (Roweis & Saul, 2000), graph for label propagation based on linear neighborhoods (LNP) (Wang & Zhang, 2008), sparse representation-based graphs (Lu, Zhou, Tan, Shang, & Zhou, 2012; Yan & Wang, 2009), and sparse probability graph (SPG) (He, Zheng, Hu, & Kong, 2011).

Sparse representation-based graph is motivated by that each datum can be reconstructed by the sparse linear superposition of other data points (Cheng, Yang, Yan, Fu, & Huang, 2010) and the sparse reconstruction coefficients are derived by solving an ℓ_1-norm regularized least square optimization problem. Unlike sparse representation which enforces the representation coefficients to be sparse, the recently proposed LRR can obtain a low-rank coefficient by solving a rank minimization problem. LRR has been widely used for various applications such as subspace segmentation (Liu, Lin, & Yu, 2010; Luo, Nie, Ding, & Huang, 2011), face recognition (Chen, Wei, & Wang, 2012) and multitask learning (Chen, Zhou, & Ye, 2011). The graph constructed by LRR can be used for many learning tasks such as spectral clustering (Liu et al., 2010) and SSL (Yang, Wang, Wang, Han, & Jiao, 2013). Several improved models have been proposed to alleviate the drawbacks of the original LRR algorithm on SSL. Non-negative low-rank and sparse (NNLRS) graph (Zhuang, Gao, Lin, Ma, Zhang and Yu, 2012) was proposed by imposing the non-negative and sparse constraints on the low-rank representation coefficient. Zheng et al. presented an algorithm to construct the graph based on low-rank representation with local constraint (LRRLC) (Zheng, Zhang, Jia, Zhao, Guo, Fu and Yu, 2013) in which the local structure is preserved by a locally constrained regularization and the global structure is preserved by LRR. A graph regularization term was added on the LRR objective and the graph regularized low-rank representation (GLRR) model was formulated for destriping of hyperspectral images in Lu, Wang, and Yuan (2013).

Recently, researchers have considered the case when data is drawn from sampling a probability distribution that has support on or near a submanifold of an ambient space. Here, a d-dimensional submanifold of an Euclidean space $\mathbb{R}^M$ is a subset $\mathcal{M}^d \subset \mathbb{R}^M$, which locally looks like a flat d-dimensional Euclidean space (Lee, 2012). It has been shown that learning performance can be significantly enhanced if the underlying manifold structure can be properly identified (Cai, He, & Han, 2011; Cai, He, Han, & Huang, 2011; Zheng, Bu, Chen, Wang, Zhang, Qiu and Cai, 2011).

Motivated by the recent progress on LRR and manifold learning, we propose a novel manifold low-rank representation model to build graph for semi-supervised classification. The basic motivation behind MLRR is to explicitly combine the global and local geometrical structure of data together in graph construction. In MLRR, the global structure is considered by the low-rank property and the local structure is emphasized by the manifold identification. Different from LRRLC and GLRR, which identify the manifold based on the Euclidean distance between data pairs, MLRR adopts the geometric sparsity idea (Elhamifar & Vidal, 2011) to approximately seek the tangent space of each data point. Here the multiple manifolds underlying the data set are assumed to be composed of many local tangent spaces (Zhang & Zha, 2004). We incorporate a regularizer into the LRR objective, aiming at enforcing the low-rank coefficients to preserve the identified manifold structure of data. Similar to NNLRS (Zhuang et al., 2012), we also constrain the representation coefficients to be sparse and non-negative. By properly identifying the manifold structure, MLRR can obtain excellent experimental results in comparison with several LRR variants and other state-of-the-art approaches.

The remainder of this paper is organized as follows. We review the original LRR, several related LRR variants and optimization method in Section 2. In Section 3, we present the formulation of proposed manifold low-rank representation model and its implementation by linearized alternating direction method with adaptive penalty (LADMAP) method (Lin, Liu, & Su, 2011). Experiments on three widely used face data sets and one voice data set to evaluate the performance of MLRR are illustrated in Section 4. Section 5 concludes the paper.

2. Related work

In this section, we review the following three parts: LRR model as well as its several variants, the LADMAP method (Lin et al., 2011) which is often employed to implement the LRR model, and the semi-supervised classification framework used in this paper.

2.1. Low-rank representation and its several variants

Let $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n] \in \mathbb{R}^{d\times n}$ be a set of n samples in the d-dimensional space. Low-rank representations aim at representing each sample by a linear combination of the bases in $\mathbf{A} = [\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_m] \in \mathbb{R}^{d\times m}$ as $\mathbf{X} = \mathbf{AZ}$, where $\mathbf{Z} = [\mathbf{z}_1, \mathbf{z}_2, \ldots, \mathbf{z}_n]$ is the matrix with each $\mathbf{z}_i$ being the representation coefficient of sample $\mathbf{x}_i$. Each element in $\mathbf{z}_i$ can be seen as the contribution to the reconstruction of $\mathbf{x}_i$ with $\mathbf{A}$ as the basis. LRR seeks the lowest-rank solution by solving the following optimization problem (Liu et al., 2010)

$$\min_{\mathbf{Z}} \operatorname{rank}(\mathbf{Z}), \quad s.t.\ \mathbf{X} = \mathbf{AZ}. \tag{1}$$

Due to the NP-hard nature of the *rank* function, the above optimization problem can be relaxed to the following convex optimization problem (Candès, Li, Ma and Wright, 2011)

$$\min_{\mathbf{Z}} \|\mathbf{Z}\|_*, \quad s.t.\ \mathbf{X} = \mathbf{AZ}, \tag{2}$$

where $\|\cdot\|_*$ denotes the trace norm of a matrix (Cai, Candès, & Shen, 2010), i.e., the sum of its singular values. Considering the fact that samples are usually noisy or even grossly corrupted, a more reasonable objective for LRR can be expressed as

$$\min_{\mathbf{Z},\mathbf{E}} \|\mathbf{Z}\|_* + \lambda\|\mathbf{E}\|_{2,1}, \quad s.t.\ \mathbf{X} = \mathbf{AZ} + \mathbf{E}, \tag{3}$$

where the $\ell_{2,1}$-norm is defined as $\|\mathbf{E}\|_{2,1} = \sum_{j=1}^{n}\sqrt{\sum_{i=1}^{d} e_{ij}^2}$ and parameter λ is used to balance the effect of low-rank term and error term. Some existing studies also used the ℓ_1-norm to measure the error term (Liu & Yan, 2012; Okutomi, Yan, Sugimoto, Liu and Zheng, 2012; Peng, Wang, Wang, & Lu, 2013) while the $\ell_{2,1}$-norm is used in this paper. The optimal solution $\mathbf{Z}^*$ to problem (3) can be obtained via the inexact augmented Lagrange multiplier (ALM) method (Lin, Chen, & Ma, 2010).

As described in Liu et al. (2010), LRR jointly obtains the representation of all the data under a global low-rank constraint, and thus is good at capturing the global structure. Moreover, since each sample can be used to represent itself, there always exist feasible solutions even when the data sampling is insufficient, which is different from sparse representation. These properties make LRR-graph a good candidate for various learning tasks. Below are several recently proposed LRR variants for graph based SSL.