Knowledge-Based Systems xxx (2014) xxx–xxx

Contents lists available at ScienceDirect

Knowledge-Based Systems

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

Semi-supervised classification via kernel low-rank representation graph

Shuyuan Yang *, Zhixi Feng, Yu Ren, Hongying Liu, Licheng Jiao

Key Lab of Intelligent Perception and Image Understanding of Ministry of Education, Department of Electrical Engineering, Xidian University, Xi'an 710071, China

ARTICLE INFO

Article history:
Received 21 June 2013
Received in revised form 29 May 2014
Accepted 8 June 2014
Available online xxxx

Keywords:
Low-rank representation
Kernel trick
Sparse graphs
Semi-supervised learning
Pairwise constraints

ABSTRACT

Sparse Representation based Graphs (SRGs) have attracted increasing interests in very recent years. However, for lacking global constraints on solutions to sparse representation, SRGs cannot accurately reveal data structure when data are grossly corrupted. In this paper, in order to achieve robust classification of wide range of datasets when only a small number of labeled samples are available, we advance a new semi-supervised kernel low-rank representation graph (SKLRG), by combining low-rank representation (LRR) with graphs and kernel trick. A kernel projection is first learned to find high-dimensional space where data have possible low-rank structure. Then a low-rank representation of the projected data is calculated from which we can derive a SKLRG matrix to evaluate data affinity and classify corrupted patterns. The proposed SKLRG can naturally reveal the relationship among data in the projected space, and can capture the global structure of complex data and implements more robust subspace segmentation. Moreover, connected weights of SKLRG are refined by pairwise constrains where label information is explored to further improve the classification results. Some experiments are taken on some benchmark datasets and Synthetic Aperture Radar (SAR) images that are corrupted by speckle noise. The results show that the proposed SKLRG can achieve better performance than its counterparts when there are only a small number of labeled samples.

1. Introduction

Semi-supervised classification using both labeled and unlabeled samples has attracted considerable attention in recent years [1–13]. It utilizes both labeled samples and the information conveyed by the unlabeled samples to boost the algorithmic performance, so making accurate classification from small labeled set possible. Graph-based models are one of the most popular semi-supervised approaches, with each sample spreads the information of labels to its neighbors in the graph [14–16]. Their basic idea is to construct a graph connecting similar samples to extract class-instance pairs from both labeled and unlabeled data. Graphs encode the intuition that similar samples have similar labels, and a graph regularization term is derived to penalize any label assignment where two nodes connected by a highly weighted edge are assigned different labels. Nowadays graph based semi-supervised classifier has proved to be able to identify classes of arbitrary distributions in many pattern recognition tasks [17–21]. In this modeling scenario, the relationship between labeled and unlabeled data remarkably impacts the construction and performance of classifiers, so constructing representative graphs to reveal the essential data structure is critical. Some commonly used graphs have been well investigated, such as k Nearest Neighbors (KNN) [22], Locally Linear Embedding (LLE) neighbors [23], ε-*ball* neighbors (ε-*ball*) [24] and so on. However, most of these graphs adopt the cluster assumption and rely on pair-wise Euclidean distances to represent local data structures, which will be sensitive to data corruption.

Recently high-dimensional statistics and optimization have brought many new tools for constructing more informative graphs. In very recent years, Sparse Representation based Graphs (SRG) attract increasing interests of researches in machine learning and image processing [25,26], where a sparse and datum-adaptive l_1-graph over a dataset is derived by encoding each datum as a sparse representation of the remaining samples [27], and automatically selecting the most informative neighbors for each datum. Although SRGs have proved to outperform some traditional graphs in practice, they assume that each sample can be individually coded by a sparse linear superposition of other samples, which emphasizes on describing the relationship of local similarity of examples and lacks global constraints on their solutions. Consequently SRGs may be inaccurate at capturing the global structures of data, especially when data are grossly corrupted.

Different with sparse representation, the recently developed low-rank representation (LRR) can find the lowest rank representation of all data jointly, to reveal the global structure of data and

* Corresponding author. Tel.: +86 029 88204298.
E-mail address: syyang@xidian.edu.cn (S. Yang).

http://dx.doi.org/10.1016/j.knosys.2014.06.007
0950-7051/

Please cite this article in press as: S. Yang et al., Semi-supervised classification via kernel low-rank representation graph, Knowl. Based Syst. (2014), http://dx.doi.org/10.1016/j.knosys.2014.06.007

correctly preserve the membership of samples that belong to the same class when there are noises [28]. The so-obtained LRR Graphs (LRRGs) can capture the global mixture of subspaces structure by the low rankness, hence is both generative and discriminative for semi-supervised classification [29,30].

A comparison of SRG and LRRG is shown in Fig. 1, where Fig. 1(a) plots the sparse representation of noisy data that consist of three classes of noisy samples, and Fig. 1(b) plots the low-rank representation of data. From them we can see that because of the existence of noises, sparse coefficients cannot accurately reveal the affinity of samples. However, when treating with corrupted data, the corruption will largely increase the rank of $\mathbf{Z}$. So the constructed LRRG is robust to noises and outliers because the low-rank criterion will enforce to correct corruptions.

Although LRRGs explore the global structure of data via the low-rankness [31–39], they have limitations in the sense that the data distribution we are trying to model is typically complex, and does not follow a strict low-rank model but a nonlinear low-rank one. Therefore, LRRGs will fail to segment data with a nonlinear union of subspaces structure. This can partially explain the failure of LRRGs works bad on some benchmark datasets (for example, some data in University of California, Irvine dataset) and some real-world data [28]. In this paper, in order to achieve robust classification of a wide range of datasets when only a small number of samples are available, we explore the global structure in the projection domain via the graph kernel, and present a new semi-supervised kernel low-rank representation graph (SKLRG). A kernel projection is first learned to find high-dimensional space where data have possible low-rank structure. Then a low-rank representation of the projected data is calculated from which we can derive an affinity matrix to reveal the complex data structure. Because LRR casts the global constraint on its solution, SKLRG can naturally reveal the relationship among data in the projected space, and can capture the global structure of data and implement more robust subspace segmentation from corrupted samples. Moreover, piecewise constraints of the graph cut formulation become applicable when the data are transformed implicitly by a kernel function. Therefore, the pairwise constraints are added into the graph to provide the supervision information: a must-link (ML) constraint specifies that the pair of instances should be assigned to the same cluster, and a cannot-link (CL) constraint specifies that the pair of instances should be placed into different clusters.

Compared with available graph-based semisupervised classification approaches, the proposed method has the following characteristics: (1) It extends the low-rankness assumption of data structure to the projection space, to formulate a new kernel LRR graph from which an accurate classification of a wider range of datasets can be obtained; (2) Pairwise constraints are added into the new graph to further improve the classification accuracy. Some experiments are taken on some benchmark datasets and Synthetic Aperture Radar (SAR) images that are corrupted by speckle noise due to random interference of electromagnetic waves, and the result show that SKLRG can achieve better performance than other related methods when only a small number of labeled samples are available.

The rest of this paper is organized as follows: In Section 2, the LLR graph, kernel LRR graph and the pairwise constraints refined kernel LRR graph are discussed. The proposed SKLRG is depicted in detail in Section 3. The experimental results and our analysis are presented in Section 4. Finally, we conclude our work in Section 5.

2. Kernel low rank representation graph with pairwise constraints

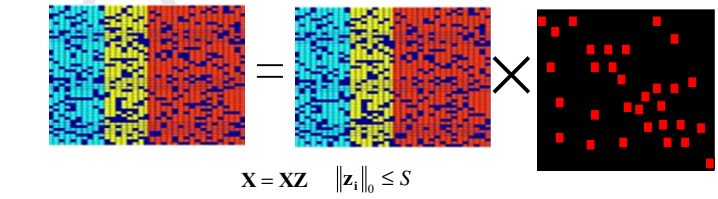
In this section, the low-rank representation and low-rank representation graph are first introduced. Then we advance a new kernel low-rank representation, from which we derive a kernel low-rank representation graph. Finally the kernel low-rank representation graph with pairwise constraints will be discussed.

2.1. Low rank representation graph (LRRG)

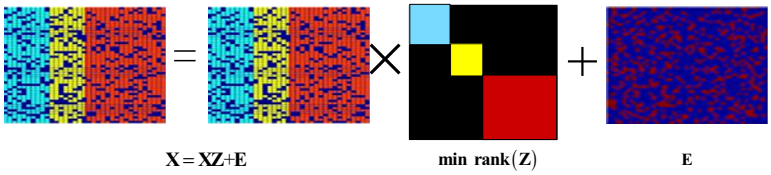
Given a set of samples $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n]$, with each datum $\mathbf{x}_i \in R^d (i = 1, ..., n)$, and assume that $\mathbf{X}$ can be represented as a linear combination of the columns (or atoms) in a dictionary $\mathbf{A}$, LRR aims to jointly find the lowest rank representation of all data: $\mathbf{X} = \mathbf{AZ} + \mathbf{E}$, where $\mathbf{Z}$ is a low-rank matrix and $\mathbf{E}$ is the noise. In our work, we use $\mathbf{X}$ itself as the dictionary for self-expressive. Therefore the low rank self-expressive representation of $\mathbf{X}$ can then be formulated as [28],

$$\begin{cases} \min\limits_{\mathbf{Z},\mathbf{E}} & rank(\mathbf{Z}) + \lambda \|\mathbf{E}\|_{2,1} \\ s.t. & \mathbf{X} = \mathbf{XZ} + \mathbf{E} \end{cases} \tag{1}$$

where $\mathbf{Z} = [\mathbf{z}_1, \mathbf{z}_2, \ldots \mathbf{z}_n]$ is the low-rank coefficient matrix, and $\lambda > 0$ is a Lagrange parameter. The l_2/l_1-norm of the matrix $\mathbf{E}$, $\|\mathbf{E}\|_{2,1}$ is defined by $\|\mathbf{E}\|_{2,1} = \sum_j \sqrt{\sum_i (\mathbf{E})_{ij}^2}$. Given a set of data drawn from a union of multiple subspaces, its low-rank representation coefficients can be used to define the affinities of an undirected graph,



(a) Sparse representation



(b) Low-rank representation

Fig. 1. Comparision of sparse representation with low-rank representation.

Please cite this article in press as: S. Yang et al., Semi-supervised classification via kernel low-rank representation graph, Knowl. Based Syst. (2014), http://dx.doi.org/10.1016/j.knosys.2014.06.007