Pattern Recognition 81 (2018) 81–94

Contents lists available at ScienceDirect

Pattern Recognition

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

Spatial and class structure regularized sparse representation graph for semi-supervised hyperspectral image classification

Yuanjie Shao[a], Nong Sang[a,*], Changxin Gao[a], Li Ma[b]

[a] National Key Laboratory of Science and Technology on Multispectral Information Processing, School of Automation, Huazhong University of Science and Technology, Wuhan 430074, China
[b] Faculty of Mechanical and Electronic Information, China University of Geosciences, Wuhan 430074, China

ARTICLE INFO

Article history:
Received 25 April 2017
Revised 17 January 2018
Accepted 27 March 2018
Available online 27 March 2018

Keywords:
Spatial regularization
Probabilistic class structure
Sparse representation (SR)
Semi-supervised learning (SSL)
Hyperspectral image (HSI) classification

ABSTRACT

Constructing a good graph that can capture intrinsic data structures is critical for graph-based semi-supervised learning methods, which are widely applied for hyperspectral image (HSI) classification with small amount of labeled samples.

Among the existing graph construction methods, sparse representation (SR)-based methods have shown impressive performance on semi-supervised HSI classification tasks. However, most SR-based algorithms fail to consider the rich spatial information of HSI, which has been shown beneficial for classification tasks.

In this paper, we propose a spatial and class structure regularized sparse representation (SCSSR) graph for semi-supervised HSI classification. Specifically, spatial information has been incorporated into SR model via the graph Laplacian regularization, it assumes that the spatial neighbors should have similar representation coefficients, the obtained coefficient matrix thus can reflect the similarity between samples more accurately. Besides, we also incorporate probabilistic class structure, which implies the probabilistic relationship between each sample and each class, into SR model to further improve discriminability of graph.

The experimental results on Hyperion and AVIRIS hyperspectral data show that our method outperforms state of the art methods.

1. Introduction

For supervised hyperspectral image (HSI) classification [1–5], we often lack sufficient labeled data which are usually costly and time-consuming to obtain. On the other hand, a large number of unlabeled data can be acquired effortlessly. Semi-supervised learning (SSL), which can make use of rich unlabeled samples, has recently been proposed to solve above problem [6,7]. Many SSL methods have been employed for HSI classification, they can be coarsely divided into four categories: (1) Expectation-maximization with generative mixture models [8]; (2) co-training based algorithm [9]; (3) Low-density separation algorithms, such as transductive support vector machines [10]; (4) Graph-based methods [11–16]. Among current SSL methods, graph-based SSL methods are particularly appealing since they have elegant mathematical formulation and can obtain a close-form solution [6,11–13,17–19].

The heart of graph-based SSL is to construct a good graph that can capture the intrinsic data structures. Graph-based SSL methods generally involve two steps. First, a graph is constructed where nodes are all samples (including labeled and unlabeled samples), and the edge weights denote the similarity between pairwise data points. Then, label information of the labeled data is propagated to the unlabeled data through the graph. Though many different objective functions have been used in graph-based SSL approaches to characterize the propagation process, most of them try to comply with the so-called cluster assumption [20], i.e., samples on the same manifold or structure are likely to share the same label. In general, the underlying manifold is approximatively modeled by a graph constructed from all samples. Therefore, graph construction is critical for graph-based SSL methods [18,21].

Many efforts have been made to learn an effective graph from data points for graph-based SSL. Traditional methods such as *k*-nearest neighbors (*k*NN) [6,11,17,18,22,23], nonnegative local lin-

* Corresponding author.
E-mail addresses: shaoyuanjie@hust.edu.cn (Y. Shao), nsang@hust.edu.cn (N. Sang), cgao@hust.edu.cn (C. Gao), mali@cug.edu.cn (L. Ma).

https://doi.org/10.1016/j.patcog.2018.03.027
0031-3203/

ear reconstruction (LLR) [19] and Local Manifold Learning (LML) [13] mainly employ Euclidean distance to build graph, and can only capture the local data structure. As a result, these methods are very sensitive to the data noise and errors. Moreover, traditional methods always have some fixed parameters which required tuning manually in graph construction, thus cannot produce data-adaptive neighborhoods.

In recent years, motivated by the advance in the area of compressed sensing [24–26], several sparse representation (SR)-based methods have been proposed to construct graphs for graph-based SSL [12,27–29]. Different from traditional methods, SR-based graphs express each sample as a sparse linear combination of all other data samples, thus are able to exploit the global structures of data. Moreover, by solving an ℓ_1 optimization problem, these methods can obtain the graph adjacency structure and weights automatically and simultaneously. Specifically, Cheng et al. [27] and Yan and Wang [28] both propose a ℓ_1-graph structure, which measures the similarity among data points by encoding each sample as a ℓ_1 sparse linear combination of the other samples. Gu and Feng [12] apply ℓ_1 graph to HSI classification. Since nonnegative characteristic is helpful for graph learning, He et al. [29] propose a sparse probability graph (SPG) by enforcing the representation coefficients to be nonnegative. To deal with noise data, Wang et al. [30] propose a modified version of SR, which is provably effective in identifying the underlying subspaces even with noisy data . Additionally, many other graphs have been developed based on SR in recent years. For example, Graph regularized sparse representation [31–33] use the graph Laplacian regularization to improve the quality of SR graph. By combining SR and low-rank representation (LRR) [34,35], Zhuang et al. [36,37] propose a nonnegative low-rank and sparse (NNLRS) model. [38,39] introduce a discriminant SR (DSR) graph structure for HSI classification, the weights of which are the product of SR weights and the class similarity weights, which are obtained by estimating the probability of two samples belonging to the same class. Recently, Shao et al. [15] propose a probabilistic class structure regularized sparse representation (PCSSR) graph construction approach for semi-supervised HSI classification, which incorporates the probabilistic class structure of the data into SR to make the representation coefficients to preserve the class structure of the data samples, and thus enhances the discriminability of graph.

While the above SR-based graphs have shown impressive performance on semi-supervised classification tasks, they all fail to consider the rich spatial information of the HSI. As we all know, spatially neighboring pixels in an HSI are likely to belong to the same class [40], and this assumption has been successfully applied in many problems [41–46]. However, these methods only consider the hypothesis of spatial consistency, but not utilize the dissimilarity relationship among samples. In the SR procedure, SR may select samples with different classes to represent target samples in the presence of dependent subspaces or data errors [47], making representation less discriminative. Therefore, in the graph construction, not only do we expect the consistency of representation coefficients of the samples in the spatial neighborhood, but also hope to penalize the samples with different class distribution (i.e, assigning smaller weights), so that the learned graph is more informative and discriminative.

Inspired by the above insights, in this paper, we present a novel graph construction method called spatial and class structure constrained sparse representation (SCSSR) approach for graph-based SSL. Specifically, spatial information has been incorporated into SR model via the graph Laplacian regularization, it assumes that the spatial neighbors have similar representation coefficients and thus obtain a more accurate coefficient matrix. On the other hand, previous work [15] has shown that probabilistic class structure can facilitate the SR and improve discriminability of graph, so we also incorporate probabilistic class structure into SR model to penalize the samples with different class distribution, and thus to further improve the performance of the graph. The SCSSR objective function can be solved by the recently developed alternating direction methods with adaptive penalty (ADMAP) algorithm [48]. The experimental results on six hyperspectral data sets show that our SCSSR method outperforms the state of the art algorithms.

We organize the rest of the paper as follows: In Section 2, we will introduce the related works, including the work related to the graph-based SSL algorithm, the work related to SR graph. In Section 3 we will detail the construction of our SCSSR graph. Experimental results and analysis will be discussed in Section 4. We will conclude our work and propose future work in Section 5.

2. Related works

In this section, we review the following two parts: graph-based semi-supervised classification framework used in the paper, and the sparse representation-based graphs.

2.1. Graph-based semi-supervised classification framework

Denote $\mathbf{Y} = [\mathbf{Y}_l, \mathbf{Y}_u]^T \in R^{n\times c}$ as a initial label matrix, where l and u represent the number of labeled samples and unlabeled samples, respectively, $\mathbf{Y}_{ij} = 1$ if the label of sample $\mathbf{x_i}$ belongs to class j for $j \in \{1, 2, \ldots, c\}$ and $\mathbf{Y}_{ij} = 0$ otherwise, c denotes the number of classes. Generally, Graph-based SSL methods obtain the final label matrix $\mathbf{F} = [\mathbf{F}_l, \mathbf{F}_u]^T \in R^{n\times c}$ by minimizing following problems [6]:

$$\min_{\mathbf{F}\in R^{n\times c}} Tr(\mathbf{F}^T\mathbf{L_W}\mathbf{F}) \quad s.t \quad \mathbf{F}_l = \mathbf{Y}_l \tag{1}$$

where $\mathbf{L_W} = \mathbf{D_W} - \mathbf{W}$ is the graph Laplacian matrix, in which $\mathbf{D_W}$ is the corresponding diagonal degree matrix with $\mathbf{D_W}_{ii} = \sum_j \mathbf{W}_{ij}$, $\mathbf{W}$ is learned affinity matrix.

By partitioning the matrix $\mathbf{L_W}$ into four blocks based on labeled and unlabeled nodes, we can obtain the closed-form solution of problem (1):

$$\mathbf{F}_u = -{\mathbf{L_W}_{uu}}^{-1}\mathbf{L_W}_{ul}\mathbf{Y}_l \tag{2}$$

2.2. Sparse representation-based graphs

The performance of the graph-based SSL depends heavily on the affinity matrix, and recent years have witnessed significant advances in graph construction. Since SR-based methods have shown impressive performance on semi-supervised HSI classification tasks, we focus our discussion on recent SR-based graphs.

Given a set of n samples $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n] \in R^{d\times n}$. SR seeks the sparsest representation among all samples that encoding each sample as the linear combination of the samples themselves. It proposes to solve the following ℓ_1 optimization problem [27,28]:

$$\begin{aligned} &\mathbf{W} = \text{argmin}\|\mathbf{W}\|_1 \\ &s.t.\ \mathbf{X} = \mathbf{X}\mathbf{W},\ \text{diag}(\mathbf{W}) = 0,\ \mathbf{W} \geq 0 \end{aligned} \tag{3}$$

where $\mathbf{W} = [\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_n]$ is coefficient matrix with each $\mathbf{w}_i$ being the representation coefficients of $\mathbf{x}_i$. The constraint $\text{diag}(\mathbf{W}) = 0$ eliminates the trivial solution of self-representation. The coefficients reflect the relationship among data samples, there existing an edge between $\mathbf{x}_i$ and $\mathbf{x}_j$ when $\mathbf{w}_{ij} \neq 0$, and the edge weight is $\mathbf{w}_{ij}$.

In order to handle with noise or corrupted data, [30] extends the above optimization problem to the following minimization problem: