Pattern Recognition 63 (2017) 102–114

Contents lists available at ScienceDirect

Pattern Recognition

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

Probabilistic 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 15 March 2016
Received in revised form
10 August 2016
Accepted 19 September 2016
Available online 21 September 2016

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

ABSTRACT

Graph-based semi-supervised learning (SSL), which performs well in hyperspectral image classification with a small amount of labeled samples, has drawn a lot of attention in the past few years. The key step of graph-based SSL is to construct a good graph to represent original data structures. Among the existing graph construction methods, sparse representation (SR) based methods have shown impressive performance on graph-based SSL. However, most SR based methods fail to take into consideration the class structure of data. In SSL, we can obtain a probabilistic class structure, which implies the probabilistic relationship between each sample and each class, of the whole data by utilizing a small amount of labeled samples. Such class structure information can help SR model to yield a more discriminative coefficients, which motivates us to exploit this class structure information in order to learn a discriminative graph. In this paper, we present a discriminative graph construction method called probabilistic class structure regularized sparse representation (PCSSR) approach, by incorporating the class structure information into the SR model, PCSSR can learn a discriminative graph from the data. A class structure regularization is developed to make use of the probabilistic class structure, and therefore to improve the discriminability of the graph. We formulate our problem as a constrained sparsity minimization problem and solve it by the alternating direction method with adaptive penalty (ADMAP). The experimental results on Hyperion and AVIRIS hyperspectral data show that our method outperforms state of the art.

1. Introduction

Hyperspectral image (HSI) data contains high-resolution spectral information on land covers, which is attractive for discriminating the subtle differences between classes with similar spectral signatures. Therefore, it has been widely applied to classify land covers [1–4]. However, HSI classification often faces the issue of limited number of labeled data, which are often costly, effortful and time-consuming to label. On the other hand, we can obtain a large number of unlabeled data effortlessly. Semi-supervised learning (SSL), which can utilize both small amount of labeled instances and abundant yet unlabeled samples, has recently been proposed to deal with this issue [5]. Several SSL methods have been applied to the classification of remote sensing and hyperspectral image, which can be divided into three classes: (1) Generative models such as expectation-maximization algorithms with finite-mixture models, which have been applied for remote sensing image classification [6]; (2) Low-density separation algorithms which maximize the margin for labeled and unlabeled data simultaneously, such as transductive SVM, which have been successfully used in HSI classification [7]; (3) Graph-based methods [8–10]. This paper focuses on graph-based SSL methods, which have drawn a lot of attention because they have elegant mathematical formulation and can obtain a close-form solution [5,11–13,8–10].

The key step of graph-based SSL is to construct a good graph to represent original data structures. Graph-based SSL methods first construct a graph where nodes are all the data samples, and the edge weights denote the similarity between pairwise data points. Then the label information of the labeled data can be effectively propagated to the unlabeled data through the graph. These methods mostly characterize the propagation process via a graph laplacian regularizer. In spite of many different forms used in graph-based SSL approaches, most regularizers try to comply with

* Corresponding author at: National Key Laboratory of Science and Technology on Multispectral Information Processing, School of Automation, Huazhong University of Science and Technology, Wuhan 430074, China.
E-mail addresses: shaoyuanjie1@gmail.com (Y. Shao), nsang@hust.edu.cn (N. Sang), cgao@hust.edu.cn (C. Gao), maryparisster@gmail.com (L. Ma).

http://dx.doi.org/10.1016/j.patcog.2016.09.011
0031-3203/

the so-called cluster assumption [14], which supposes that the data on the same manifold or structure are likely to have the same class. In most methods, the underlying manifold is approximatively modeled by the construction of a graph from both labeled and unlabeled samples. Therefore, constructing a good graph to represent underlying data manifold is critical for graph-based SSL approaches [12,15,16].

To learn an effective graph from data points, many techniques have been proposed for graph-based SSL. Traditional methods often utilize k-nearest neighbors (kNN) or ε-ball neighborhood to determine the graph adjacency structure, and then employ kinds of weight calculation methods, e.g., gaussian kernel (GK) function [12,5,17,11,18,8] and nonnegative local linear reconstruction (LLR) [13] coefficients to measure the similarity among data samples. Ma et al. [10] propose a new family of graph weights based on Local Manifold Learning (LML) for HSI classification. The shared problems of these graph construction methods are that they all have some fixed parameters which required tuning manually, and the graph structures are also very sensitive to the parameter variance and data noise.

Recently, benefiting from the development of compressed sensing [19,20], several sparse representation (SR) based graphs have been proposed for graph-based SSL [21–23,9]. For example, both Cheng et al. [21] and Yan et al. [22] propose a ℓ_1-graph structure based on SR.The latent philosophy is that each sample can be encoded as a sparse linear combination of the remaining samples via solving an ℓ_1 optimization problem, and the sparse coefficients could represent the similarity between two samples. By enforcing the representation coefficients to be nonnegative so that it can be directly applied as graph weights, He et al. [23] provide a sparse probability graph (SPG). In order to handle with noise or corrupted data, [24] proposes a modified version of SR, which can correctly identify the underlying subspaces even with noisy data. Besides, Gu and Feng [9] also apply SR graph to HSI classification. In contrast to aforementioned graphs (i.e., GK graph, LLR graph, LLE graph), SR-based graphs can obtain the adjacency relationship and the weights automatically and simultaneously, and has natural discriminating power. Many SR variants have also been developed. Graph regularized sparse representation (GRSR) [25–27] considers the local manifold structure of the data, and the obtained sparse representations vary smoothly along the geodesics of the data manifold. Zhuang et al. [28,29] propose a nonnegative low-rank and sparse (NNLRS) model, by combining SR and low-rank representation (LRR) [30,31]. The so obtained graph can capture both the global cluster structure and local linear structure of the data.

As discussed above, the SR based graphs have shown impressive performance on semi-supervised classification tasks. However, most of those SR-based graph construction methods fail to take the class structure of the data into account. In SSL, we can obtain a probabilistic class structure, which implies the probabilistic relationship between each sample and each class, of the whole data by utilizing partial label information. Such class structure information can help SR model to yield a more discriminative coefficients, e.g., if two samples have different class distribution, which means they belong to the same class with low probability, we would like to assign small weight to them. Thus it is beneficial for exploiting this class structure information in order to learn a discriminative graph.

Motivated by above insights, in this paper, we present a discriminative graph construction method called probabilistic class structure regularized sparse representation (PCSSR) approach for graph-based SSL. Specifically, a class structure regularizer is developed to incorporate the probabilistic class structure information into SR model. This regularizer can enable us to learn a discriminative graph, which benefits the graph-based semi-supervised classification task. The PCSSR objective function can be solved by the recently developed alternating direction with adaptive penalty (ADMAP) algorithm [32]. The experimental results on six hyperspectral data sets show that our PCSSR method outperforms the state of the art algorithms in graph construction in most cases.

Summarily, the main contributions of this paper are as follows.

(1) We provide a method to estimate the probabilistic class structure, which implies the probabilistic relationship between each data point and each class, of the whole data. Such prior information can regularize the graph construction approach and result in a more suitable graph for classification.
(2) We propose a PCSSR graph construction approach, which incorporates the class structure information of the data into the formulation of SR and thus enhances the discriminability of graph.

We organize the rest of the paper as follows: In Section 2, we will introduce related work, including the work related to the graph-based SSL algorithm, the work related to some existing graph construction methods. In Section 3 we will detail the construction of PCSSR 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 first introduce a representative graph-based SSL method: label propagation (LP) algorithm, and then discuss some existing graph construction methods.

2.1. Label propagation algorithm

In this part, we will briefly review the LP algorithm proposed in [5,17], and apply our PCSSR graph to it in this paper. LP algorithm aims to propagate labels from limited labeled samples to abundant unlabeled samples. Given a little number of labeled samples, the basic idea of LP method is to propagate the labels according to affinity relationship among samples. It assumes that similar samples should have similar labels. The solution of the algorithm is first to compute the similarities among all the data samples, and then formulate the label propagation problem as a harmonic energy minimization problem [5], which has a closed-form solution. We summarize the algorithm as follows.

Given the labeled samples $\mathbf{X}_l = [\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_l]$ and the unlabeled samples $\mathbf{X}_u = [\mathbf{x}_{l+1}, \mathbf{x}_{l+2}, \ldots, \mathbf{x}_{l+u}]$, they have c classes denoted as $C = [1, 2, \ldots, c]$. Both the labeled and unlabeled samples $\mathbf{X} = [\mathbf{X}_l, \mathbf{X}_u]$ produce a connected graph $G = (V, E)$, where the nodes V corresponding to the $n = l + u$ samples, and the edges E are represented by a similarity weight matrix $\mathbf{W} \in R^{n\times n}$. Let $\mathbf{Y} = [\mathbf{Y}_l, \mathbf{Y}_u]^T \in R^{n\times c}$ be a label matrix, where $\mathbf{Y}_{ij} = 1$ if the label of sample $\mathbf{X}_i$ belongs to class j for $j \in C$ and $\mathbf{Y}_{ij} = 0$ otherwise. The goal of label propagation is to learn a prediction function $\mathbf{F} = [\mathbf{F}_l, \mathbf{F}_u]^T \in R^{n\times c}$ by minimizing following harmonic energy function, meanwhile it constrains the prediction labels of labeled samples to be equal to true labels:

$$\min_{\mathbf{F}\in R^{n\times c}} \frac{1}{2}\sum_{i,j=1}^{N} \mathbf{W}_{ij}\left\|\mathbf{f}_i - \mathbf{f}_j\right\|^2 = Tr(\mathbf{F}^T\mathbf{L}_\mathbf{W}\mathbf{F})\ s.t\quad \mathbf{F}_l = \mathbf{Y}_l \tag{1}$$

where $\mathbf{f}_i \in R^{1\times c}$ and $\mathbf{f}_j \in R^{1\times c}$ are predicted label vector of data sample $\mathbf{x}_i$ and $\mathbf{x}_j$, $\mathbf{L}_\mathbf{W} = \mathbf{D} - \mathbf{W}$ is the graph Laplacian matrix, in which $\mathbf{D}$ is the diagonal degree matrix with $\mathbf{D}_{ii} = \sum_j \mathbf{W}_{ij}$.

Partition the matrix $\mathbf{L}_\mathbf{W}$ into four blocks based on the number