Neurocomputing 173 (2016) 332–338

Contents lists available at ScienceDirect

Neurocomputing

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

Set-based similarity learning in subspace for agricultural remote sensing classification

Yi Tang [a], Xinrong Li [b]

[a] Key Laboratory of IOT Application Technology of Universities in Yunnan Province, Department of Mathematics and Computer Science, Yunnan Minzu University, Kunming, Yunnan 650500, PR China
[b] Institute of Plant Nutrition and Natural Resources, Beijing Academy of Agriculture and Forestry Sciences, Beijing 100097, China

ARTICLE INFO

Article history:
Received 30 June 2014
Received in revised form
24 September 2014
Accepted 12 November 2014
Available online 11 September 2015

Keywords:
Agricultural remote sensing
Point-based similarity
Set-based similarity
Similarity learning
Hyperspectral image classification

ABSTRACT

Similarity between spectral lines is key in the field of agricultural sensing classification, however, the measured spectral lines mostly mislead the classification because of unexpected disturbance in application. To enhance the accuracy of classification, similarity learning is introduced into agricultural remote sensing classification. Within the framework of similarity learning, the training set is generated by pairing the labeled spectral lines which means the size of training set for learning similarity is heavily increasing. Noticed this problem, a novel spectrum-set similarity learning algorithm is reported for balancing the gain in classification and the computational burden of learning similarity. Different from traditional point-based similarity, the spectrum-set similarity measures the similarity between two spectral sets which contain some spectral lines. Following the idea, set-based training set is generated by clustering the spectral lines in the point-based training set. Experimental results have shown the effectiveness and efficiency of learning spectrum-set similarity measure for agriculture sensing classification.

1. Introduction

Distinguishing the crops on the ground is an important application of agricultural remote sensing. Hyper-spectral information of kinds of crops is recorded by sampling hundreds or even thousands of bands of continuous spectra, which contains rich discriminant information on classifying crops. Because of the discriminant information, different crops could be classified by measuring the similarity between their spectral lines. Though agricultural remote sensing classification is possible by directly using spectral information, it is more efficient and effective to modify the spectral information for classification.

It is prevalent to learn discriminant information of spectral lines via subspace learning [1–4] in the field of hyper-spectral image classification [5]. The discriminant information of original spectral lines is extracted by representing the original spectral lines in a low-dimensional space explicitly or implicitly [6]. For example, general subspace learning based methods include the linear discriminant analysis [7,8] and nonparametric weighted feature extraction [9].

Based on these subspace based methods, more efficient and effective methods are also proposed for dealing with the specific problems of agricultural remote sensing classification. For instants, regularized linear discriminant analysis [10] and refined Fisher's linear discriminant analysis [8] are introduced into for releasing the interference of the scarcity of labeled spectral lines. Taking advantages of unlabeled spectral lines, manifold subspace learning methods [11,12] and graph-based semi-supervised learning methods are also employed. The core of these methods is to design an appropriate similarity measure in a low-dimensional subspace to boost the separability of training samples.

Similarity measure is the paraphrase of the labeled information of training spectral lines. For example, the samples are similar if they share same label index, but not vice versa. Unfortunately, the similarity induced by label indices may be inconsistent with the similarity induced by a similarity measure defined in the spectral feature space. So learning a suitable similarity measure is a hot topic in the field of designing classification algorithms. Especially, the process of learning similarity in a spectral feature space is mostly a part of many nearest-neighbor-based classification algorithms. For example, nearest-neighbor-based nonparametric feature extraction method [13] emphasizes the importance of refining the similarity measure, and the similarity of matrix-based features is employed in [14].

The relation between classification and similarity learning motivates us to consider the problem of agricultural remote sensing classification via learning an appropriate similarity measure.

E-mail addresses: yitang.math@ieee.org (Y. Tang), xr0955@sina.com (X. Li).

http://dx.doi.org/10.1016/j.neucom.2014.11.100
0925-2312/

By using nearest-based classifiers as medias, learning a classifier is equivalent to learning a similarity measure which is represented by a low-rank and positive-definite matrix. It should be noticed that the size of training set gets much huger when a problem of classification turns into a problem of learning similarity measure. For example, given n labeled training samples for classification, $O(n^2)$ labeled training samples for learning similarity measure could be generated by pairing the samples for classification.

Motivated by the success of set–set distance learning [15,16], spectrum-set similarity measure is introduced in this paper for responding to the challenges of large scale problem of treating a classification as learning similarity. Different from traditional similarity learning algorithms such as [17–19], spectrum-set similarity measure defines the similarity between training subsets. Because training subsets replace training points, the size of training set will be dramatically reduced, which could improve the efficiency of learning similarity measure.

The main contribution of this paper is treating agricultural remote sensing classification as a set-based similarity learning which is helpful in balancing the effectiveness of classification and the efficiency of learning similarity measure.

The rest of this paper is organized as follows. The results about main algorithm are reported in Section 2. Experimental results have shown in Section 3. The paper is ended with a conclusion in Section 4.

2. Main algorithm

2.1. Point-based model

In this subsection, a point-based model is introduced for unveiling the core idea of classification via learning similarity measure. Let $\{(x_1,y_1),(x_2,y_2),\ldots,(x_n,y_n)\}\subseteq\mathcal{X}\times\mathcal{Y}$ be a training set containing n training samples where $\mathcal{X}$ is a spectral feature space and $\mathcal{Y}=\{1,2,\ldots,k\}$ is the label indices space. For a test sample $x\in\mathcal{X}$, the label of x could be estimated by comparing the similarity between training and test spectral features. Specifically, the estimated label $\hat{y}$ corresponding to x is defined by y_{i^*} where

$$i^*=\arg\min_{i=1,2,\ldots,n} S(x,x_i). \tag{1}$$

Here, $S(\cdot,\cdot)$ is a similarity measure function defined on $\mathcal{X}\times\mathcal{X}$.

The similarity measure function $S(\cdot,\cdot)$ is generally defined by Euclidean distance. Thus, the optimization problem (1) could be expressed as

$$i^*=\arg\min_{i=1,2,\ldots,n}\|x-x_i\|_2^2. \tag{2}$$

Though the Euclidean distance based similarity measure is popular in application, the consistency between the similarity measures induced by Euclidean distance and the labels of spectral lines is always suspicious. To enhance the consistency between both similarity measures, the original spectral features are mostly mapped into a low-dimensional subspace by a low-rank projection matrix P. That is

$$i^*=\arg\min_{i=1,2,\ldots,n}\|P(x-x_i)\|_2^2. \tag{3}$$

The best low-rank projection matrix P^* which promises the best consistency between distance based similarity and index based similarity could be learned with the help of training data. Because the labels of training samples are given, the index based similarity between any training samples (x_i,x_j) could be defined by their labels. For example

$$S_I(x_i,x_j)=\begin{cases}1 & y_i=y_j\\ -1 & y_i\neq y_j\end{cases} \tag{4}$$

where $S_I(\cdot,\cdot)$ is the index based similarity measure function.

Taking advantages of S_I, the information of class labels could be explained as similarity information. The triple set

$$T=\{(x_i,x_j,S_I(x_i,x_j))\,|\,i,j=1,2,\ldots,n\} \tag{5}$$

describes the consistency between the distance based similarity and the index based similarity where the distance based similarity could be defined as

$$S_D(x_i,x_j)=\begin{cases}1 & \|P(x_i-x_j)\|_2^2<\gamma,\\ -1 & \|P(x_i-x_j)\|_2^2\geq\gamma.\end{cases} \tag{6}$$

Here $\gamma>0$ is a threshold. The consistency between the index based similarity measure $S_I(x_i,x_j)$ and the distance based similarity measure could be measured by verifying the equation $S_I(x_i,x_j)=S_D(x_i,x_j)$.

According to the rule of empirical risk minimization [20–22], the best consistency between index based similarity and distance based similarity could be generated by the optimal low-rank projection matrix P^* which is obtained by optimizing the empirical risk function

$$P^*=\arg\min_P\sum_{1\leq i,j\leq n}L(S_I(x_i,x_j),\mathrm{sign}(\gamma-\|P(x_i-x_j)\|_2^2)), \tag{7}$$

where $L(\cdot)$ is a classification risk function [23,24] defined on real and satisfies Lipschitz condition, and $\mathrm{sign}(\cdot)$ is the sign function. Notice that

$$\|P(x_i-x_j)\|_2^2=\langle P'P,(x_i-x_j)(x_i-x_j)'\rangle_F \tag{8}$$

where P' is the transpose matrix of P and $\langle\cdot,\cdot\rangle_F$ is the Frobenius inner. Thus the problem (7) could be rewritten by using the formula (8) as

$$P^*=\arg\min_P\sum_{1\leq i,j\leq n}L(S_I(x_i,x_j),\mathrm{sign}(\gamma-\langle PP',(x_i-x_j)(x_i-x_j)'\rangle_F)). \tag{9}$$

Denote $PP'=M$. It is clear that M is a low-rank, symmetric and positive-definite matrix because P is a low-rank matrix. So the problem of learning the low-rank projection matrix P is equivalent to the problem of learning the low-rank and positive-definite matrix M, that is,

$$M^*=\arg\min_M\sum_{1\leq i,j\leq n}L(S_I(x_i,x_j),\mathrm{sign}(\gamma-\langle M,(x_i-x_j)(x_i-x_j)'\rangle_F)). \tag{10}$$

Considering the low-rank and positive-definite restrictions of symmetric matrix M, the optimization problem (10) could be expressed as its regularization version

$$M^*=\arg\min_M\sum_{1\leq i,j\leq n}L(S_I(x_i,x_j),\mathrm{sign}(\gamma-\langle M,(x_i-x_j)(x_i-x_j)'\rangle_F))+\mu\|M\|_*, \tag{11}$$

where $\mu>0$ is the regularization parameter, and $\|\cdot\|_*$ is the nuclear norm which is used as a relaxation of the low-rank restriction.

Combining the formula (3) and the learned low-rank matrix M^*, a refined classification could be generated by a refined similarity measure which means

$$\begin{aligned}i^*&=\arg\min_i\|P^*(x-x_i)\|_2^2\\&=\arg\min_i\langle M^*,(x-x_i)(x-x_i)'\rangle_F.\end{aligned} \tag{12}$$

Therefore, classification could be understand as a problem of learning similarity in a subspace which leads to a low-rank similarity matrix learning problem (11).