ISPRS Journal of Photogrammetry and Remote Sensing 105 (2015) 19–29

Contents lists available at ScienceDirect

ISPRS Journal of Photogrammetry and Remote Sensing

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

A novel semi-supervised hyperspectral image classification approach based on spatial neighborhood information and classifier combination

Kun Tan [a], Jun Hu [a], Jun Li [b], Peijun Du [c,*]

[a] Jiangsu Key Laboratory of Resources and Environment Information Engineering, China University of Mining and Technology, Xuzhou, China
[b] Guangdong Key Laboratory for Urbanization and Geo-Simulation, School of Geography and Planning, Sun Yat-Sen University, Guangzhou, China
[c] Key Laboratory for Satellite Mapping Technology and Applications of State Administration of Surveying, Mapping and Geoinformation of China, Nanjing University, Nanjing, China

ARTICLE INFO

Article history:
Received 2 November 2014
Received in revised form 11 March 2015
Accepted 13 March 2015
Available online 13 April 2015

Keywords:
Semi-supervised classification
Hyperspectral images
Spatial neighborhood information
Classifier fusion
Multinomial logistic regression (MLR)
k-nearest neighbor (KNN)

ABSTRACT

In the process of semi-supervised hyperspectral image classification, spatial neighborhood information of training samples is widely applied to solve the small sample size problem. However, the neighborhood information of unlabeled samples is usually ignored. In this paper, we propose a new algorithm for hyperspectral image semi-supervised classification in which the spatial neighborhood information is combined with classifier to enhance the classification ability in determining the class label of the selected unlabeled samples. There are two key points in this algorithm: (1) it is considered that the correct label should appear in the spatial neighborhood of unlabeled samples; (2) the combination of classifier can obtains better results. Two classifiers multinomial logistic regression (MLR) and k-nearest neighbor (KNN) are combined together in the above way to further improve the performance. The performance of the proposed approach was assessed with two real hyperspectral data sets, and the obtained results indicate that the proposed approach is effective for hyperspectral classification.

1. Introduction

Due to the special advantages of the wide spectral range, high spectral resolution, and continuous spectral curve, hyperspectral remote sensing images have been widely applied in earth observation (Wilson and Felt, 1998; Yao-hua et al., 2012). However, the development of hyperspectral remote sensing technology has also brought huge challenges: (1) high-dimensional data sets usually contain redundant information, which increases the amount of computation and leads to the Hughes phenomenon (Richards and Richards, 1999). They are also easily influenced by noise and water absorption (Bruce et al., 2002; Du et al., 2003); (2) obtaining labeled training samples is generally expensive, difficult, and time-consuming (Tan et al., 2014). In recent years, novel discriminative approaches, such as artificial immune network (AIN) (Zhong and Zhang, 2012; Zhong et al., 2006), support vector machines (SVM) (Camps-Valls et al., 2008, 2006; Plaza et al., 2009; Tan and Du, 2008), deoxyribonucleic acid (DNA) computation (Jiao et al., 2012), extreme learning (Samat et al., 2014), minimum spanning forest (MSF) (Tarabalka et al., 2010) etc., have been proposed to tackle the aforementioned difficulties for remote sensing image processing tasks. However, it is often difficult for a traditional classifier to offer satisfactory performance in hyperspectral image classification especially with limited small training set. This observation has fostered the idea of semi-supervised learning, which adds unlabeled samples to the training set, without significant cost, to improve the capability of the classifier (Shahshahani and Landgrebe, 1994). In general, semi-supervised learning consists of five different models: generative models (Chapelle et al., 2006; Jin et al., 2013), graph-based methods(Zha et al., 2009; Zhu and Lafferty, 2005), transductive support vector machines (TSVMs) (Joachims, 1999; Tong and Koller, 2002), self-learning approaches (Grandvalet and Bengio, 2005; Rosenberg et al., 2005; Tuia et al., 2009), and multiview learning (Di and Crawford, 2010, 2011, 2012).

The main problems of semi-supervised learning approach are how to select the most helpful unlabeled samples and how to determine the class label of these new selected samples. In this paper breaking ties (BT) (Luo et al., 2004) method is applied to select the most useful unlabeled samples. It can greatly reduce the amount of computing time and improve the efficiency of the algorithm. At first, the label is estimated just through the initial classification map (Dópido et al., 2013). However, a small number

* Corresponding author. Tel.: +86 15905159291.
E-mail address: dupjrs@126.com (P. Du).

http://dx.doi.org/10.1016/j.isprsjprs.2015.03.006
0924-2716/

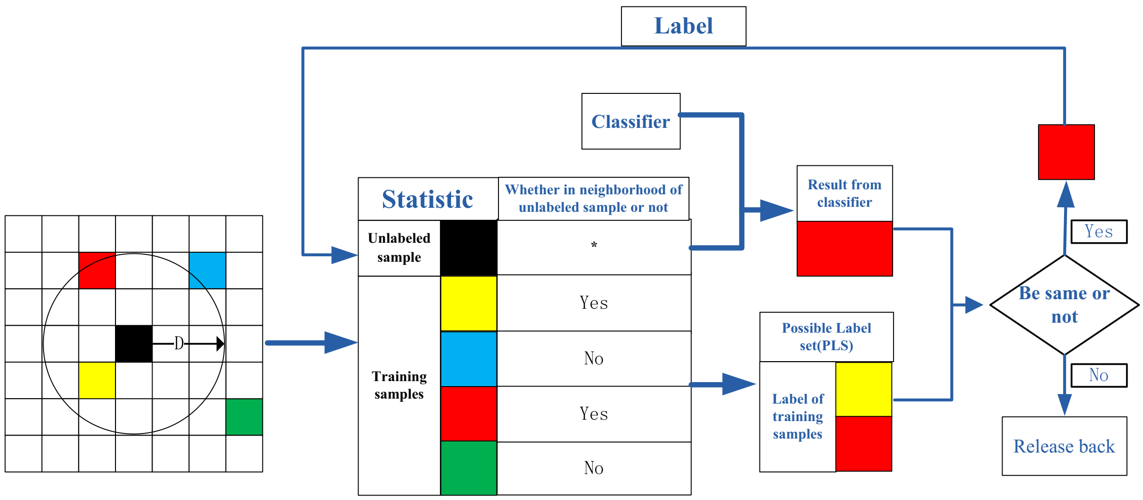

Fig. 1. The label process for unlabeled samples based on spatial neighborhood information.

of initial labeled samples, likely with poor generalization, makes the classification problem very difficult. A modified BT method, called MBT, was integrated in classification (LORSAL) and segmentation (LORSAL-MLL), resulting in two new methods with active learning, called LORSAL-AL and LORSAL-MLL-AL (Li et al., 2011). Unlabeled samples are applied to improve the estimation of the class distributions, and the obtained classification is refined by using a spatial multi-level logistic prior (Li et al., 2010). Begüm proposed a strategy that can be embedded in any AL method to identify the most informative samples and to reduce the overall cost (Demir et al., 2014). With the combination of spectral and spatial information, it is widely used in remote sensing image classification, leading to obvious improvements to the classification accuracy (Bioucas-Dias et al., 2013; Li et al., 2013). As a consequence, discriminative approaches can mitigate the curse of dimensionality because they just require smaller training sets (Di and Crawford, 2012; Tan et al., 2014).

In this paper, a novel approach is proposed to confirm the labels of unlabeled samples. This method is to reduce the difficulty of samples selection in the semi-supervised classification based on the spatial neighborhood information and classifier combination. "Spatial Neighborhood Information of Labeled samples" (SNI-L) based on 4-neighborhood or 8-neighborhood is usually applied in the semi-supervised learning process (Dópido et al., 2013; Wang et al., 2014). However, the spatial neighborhood information of the selected unlabeled samples (SNI-unL) is rarely used in the determination process of samples label. When the class of each pixel is known, all the pixels can be regarded as training samples and the label of unlabeled samples must be same with one of the 8-neighborhood pixels. When the amount of initial training samples is small, the label of unlabeled samples also should be same as one of the training samples' appearing in the neighborhood, but the nearest training sample may not be the right label. So the final label should be judged by classifier and the information of near training samples is helpful for the classifier. If the label assigned by a classifier is same as that of the training samples' appearing in the neighborhood, it could be chosen as right training samples.

At present, support vector machines (SVM) (Schölkopf and Smola, 2002), multinomial logistic regression (MLR) (Böhning, 1992), and ensemble classifiers (EC) (Du et al., 2012) are widely used. There is still a need to discuss the final label given only by one classifier. So, in order to improve the accuracy of sample selection, two classifiers multinomial logistic regression (MLR) and k-nearest neighbor (KNN) (Li et al., 2005) are combined together.

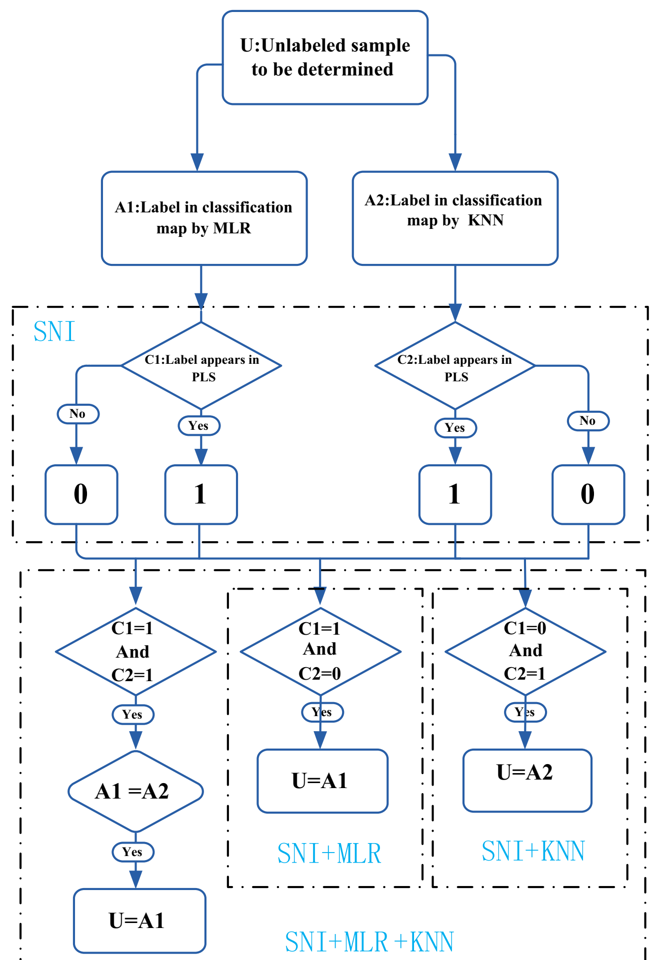

Fig. 2. The detailed process of label for the unlabeled sample.

The remainder of this paper is organized as follows. Section II describes the proposed approach for semi-supervised self-learning. Section III gives the classification results of two real hyperspectral images collected by the Airborne Visible-Infrared Imaging Spectrometer (AVIRIS) (Green et al., 1998) and the Reflective Optics Spectrographic Imaging System (ROSIS) (Benediktsson