AN ACTIVE LEARNING METHOD BASED ON SVM CLASSIFIER FOR HYPERSPECTRAL IMAGES CLASSIFICATION

Shujin Sun, Ping Zhong, Huaitie Xiao, Fang Liu, Runsheng Wang

ATR Lab. School of Electronics Science and Engineering,
National University of Defense Technology,
Changsha 410073, Hunan, China

ABSTRACT

Support vector machine (SVM) classifiers have been one of the most popularly used classifier for hyperspectral image classification. However, the amount of labeled samples available in practice is often few and the collection of the labeled set is time-consuming. To collect the most informative training set and train an effective classifier, we proposed a new active learning (AL) method with SVM classifier. The proposed AL method was based on the one-against-one multiclass strategy. In addition, to alleviate the computational burden during the AL steps, an incremental strategy for the model updating process was introduced. Experimental results on the AVIRIS Indian Pines data set demonstrated the competitive performance and better stability of the proposed AL approach with respect to the state-of-the-art AL technique, as well as the efficiency of the proposed incremental strategy during the AL process.

***Index Terms*—** Hyperspectral image classification, active learning, support vector machine

1. INTRODUCTION

With the development of the remote sensed instruments, hyperspectral image classification has been an active area of research in remote sensing community [1]. The main difficulty lies in the few labeled samples versus high dimensional features, known as the Hughes phenomenon [2]. Kernel methods such as the support vector machine (SVM) [3] have been widely used to deal effectively with the Hughes phenomenon.

Despite their excellent performances, SVMs, as any supervised classifier, rely on the quality of the labeled data used for training. However, the amount of labeled samples available in practice is often few and the collection of the labeled set is time-consuming. A well-known trend to deal with the problem is to efficiently construct small training sets with high training utility with active learning (AL) methods [4, 5].

The aim of AL is to rank the unlabeled set according to a criterion that allows us to select the most useful samples to improve the classifier, thus minimizing the number of training samples necessary to maintain discrimination capabilities as high as possible [6]. A complete review of AL can be found in [7]. The research on AL with SVM classifiers has been an hot topic in recent years, among which marginal sampling (MS) and multiclass level uncertainty (MCLU) are the two most famous AL methods in hyperspectral image classification [5]. The MS strategy samples the candidates lying within the margin of the current SVM by computing their distance to the dividing hyperplane, while MCLU considers the difference between the distance to the margin for the two most probable classes. Note that, the AL methods of MS and MCLU are both based on the one-against-all (OAA) multiclass strategy for SVM, although the one-against-one (OAO) strategy has been shown to be more suitable for practical use [8].

In this paper, we focus on design AL method with SVM classifier, and a new AL approach is proposed based on the OAO multiclass startegy. Our proposed AL method intends to evaluate the confusion of the two most likely class labels, and select the instance which has the least confidence to discriminate the two classes. In the proposed AL method, the two most likely classes are firstly identified for each instance in the unlabeled set, and then the absolute discriminant value predicted by the binary classifier for the two classes is computed. Finally, the instances characterized by the minimum absolute discriminant value are selected as the most informative samples for manual labeling. In addition, to reduce the computational cost, an incremental strategy for classifier training and prediction during the AL steps is introduced. By using the incremental strategy, the repeated training of binary classifiers could be avoided, and the computation of the discriminant function values associated is not needed either, thus reducing the computational cost.

The main contributions of this work lie in two aspects: 1) A new AL approach with SVM classifier was proposed based on the OAO multiclass strategy; and 2) An incremental strategy was introduced to efficiently implement the AL process.

The rest of this paper is organized as follows. Section 2

This work was supported by the NSF of China under Grant 61271439 and Grant 61372159, FANEDD under Grant 201243, Program for New Century Excellent Talents in University under Grant NECT-13-0164, and Hunan Provincial NSF of China under Grant 2015JJ3018.

introduces the multiclass strategy of the SVM classifier and the proposed AL approach. Section 3 illustrates the experimental results obtained on the AVIRIS Indian Pines data set. Finally, Section 4 concludes with some remarks and hints at plausible future work.

2. PROPOSED METHOD

2.1. SVM Multiclass Classification

SVMs are intrinsically binary classifiers. However, in practical image classification tasks, it often needs the simultaneous discrimination of several different classes. Generally, the multiclass extension of SVM classifiers can be done by combining several binary classifiers, among which two classical strategies OAA and OAO.

The OAA method constructs M binary classifiers where M is the number of classes. The ith classifier is trained with all of the samples in the ith class with positive labels $(+1)$ and all other examples with negative labels (-1). For a new test instance, the class with the largest discriminant function value is the predicted class label for that instance. While in the OAO multiclass strategy, it constructs $M \times (M-1)/2$ binary classifiers where each one is trained on the labeled samples from every two classes. For a new test sample, we could predict the class label by the "Majority Vote" strategy, in which the class which has the largest number of votes by all the binary classifier is set as the predicted class for the test instance.

2.2. Proposed AL Approach

Let us consider an initial training set D_L and an additional unlabeled set D_U. In order to collect the most informative training set and learn a good classifier, the AL algorithm has the task of choosing the most informative samples from D_U for manually labeling and enlarging the training set D_L iteratively. Because the inclusion of a single candidate per iteration is not optimal in remote sensing image classification tasks, the batch mode AL method that selects several candidates per iteration is often preferable.

Conventional AL methods with SVM classifier,such as marginal sampling (MS) [5] and multiclass level uncertainty (MCLU) [5] have been demonstrated effective to collect the informative training set in the classification of hyperspectral images. However, they are mainly based on the OAA multiclass strategy, and the AL methods designed with the OAO strategy are few. In [6], the marginal sampling (MS) AL strategy was generalized to be adopted by the OAO strategy. With MS method, the class w_1^i with the largest number of votes is firstly identified for each unlabeled candidate $\mathbf{x}_i$, and then the $M-1$ binary classifier associated with the class w_1^i are considered to calculate the minimum absolute value of the discriminant function f_i. Finally, the samples characterized by the minimum values of f_i are selected.

Algorithm 1 Proposed Active Learning Approach

Input:
Labeled set D_L;
Unlabeled set D_U;
Number of all classes M;
Batch size of selected instances at each iteration h;
Output:
Selected set of instances D_S;
1: Train the $M \times (M-1)/2$ binary SVM classifier with the current training set;
2: For each instance $\mathbf{x}_i$ in D_U, calculate the vote numbers $\{v_1, v_2, ..., v_M\}^i$ for all the classes, and identify the two mostly voted classes w_1^i, w_2^i for each instance;
3: Predict the discriminant function value f_i for the ith candidate using the binary classifier trained for the two classes w_1^i and w_2^i;
4: Initialize D_S to empty set, and include h samples which have the lowest absolute value of the discriminant function output $|f_i|$ into D_S from D_U;
5: The expert adds the label to each unlabeled instance in D_S, and these samples are added to the current training set D_L and removed from D_U.

In this work, we adopt OAO multiclass classification, which has shown to be more suitable for practical use than OAA strategy [8], and propose a new AL approach for SVM calssifier based on the OAO multiclass strategy. While MS method selects the instances which are the most uncertain, and the uncertainness is measured by the least confidence value of the $M-1$ binary classifier associated with the most likely class for that instance. Our proposed AL method intends to evaluate the confusion of the two most likely class labels, and select the instance which has the least confidence to discriminate the two classes. We hope that the selection of this type of unlabeled instance would help construct a useful set of training samples and train a good classifier.

Our AL approach consists of three main steps: 1) Calculate the vote number for each class for each $\mathbf{x}_i$ in the unlabeled set, and identify the two mostly voted classes w_1^i, w_2^i; 2) Calculate the minimum absolute value of the discriminant function $|f_i|$ by the binary classifier trained for the two classes w_1^i and w_2^i; 3) Select the instances characterized by the minimum values of $|f_i|$. Algorithm 1 gives the detailed description of the proposed AL method at one iteration.

In addition, to reduce the computational cost, an incremental strategy for classifier training and prediction during the AL steps is introduced. In fact, after the manual labeling step for the selected instances, the newly labeled samples often do not cover all the classes. That is to say, several binary classifiers might not need to be trained between two consecutive AL iterations. Moreover, the calculation of the discriminant function outputs for the instances in the remaining unlabeled set are unnecessary. In the incremental strategy,

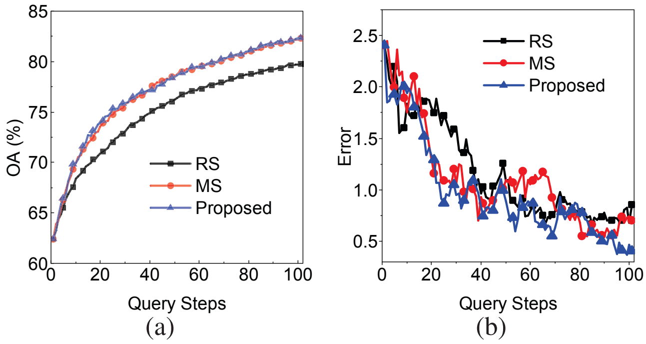


Fig. 1. Classification performance (OA) of different AL methods on the AVIRIS Indian Pines data set. a) Classification results of different AL method in function with the query steps averaged over 10 runs of the algorithms. b) Standard deviation values associated with each AL method.

a set W^* is firstly obtained by the unique classes covered in the newly selected instances after each iteration. Then, for a new iteration, only the binary classifiers for the class pairs involved in W^* are trained, and consequently the discriminant function values for the instances in current unlabeled set are not needed to be calculated either. In this way, the repeated computation for training certain binary classifiers and the corresponding discriminant function values of the unlabeled instances could be avoided in the new query step. Taking a $M-$class classification problem for example, if N classes $(N < M)$ are covered by the newly labeled samples, the retraining of $(M-N)\times(M-N-1)/2$ binary classifiers could be avoided, which may potentially reduce the computational cost during the AL iterations.

3. EXPERIMENTAL RESULTS

To evaluate the performance of the proposed AL approach, experiment was conducted on the AVIRIS Indian Pines hyperspectral data set. The image has spatial dimensions of 145 by 145 pixels, and a spatial resolution of 20 m per pixel. Twenty water absorption bands have been removed, and a 200-band image was used for the experiments. In the experiments, six classes were discarded for their small size, which results in a ten-class classification problem with 9725 labeled pixels.

Among the available labeled pixels, AL was randomly initiated with 200 pixels (20 points per class), in order to take into account enough information for all the classes. The remaining 9525 pixels were randomly split into an unlabeled candidate set with 4763 pixels and test set with 4762 pixels. In the experiments, ten samples were selected at each iteration. To reach convergence, 100 iterations needed to be executed for the data set. For experimental purpose, in the present work, the collected unlabeled candidates were labeled using the ground truth information.

To evaluate the performance of the proposed approach, two other sample selection methods were also included in

Table 1. Classification results (OA, AA, and Kappa) achieved on the Indian Pines data with the initial labeled set and the final labeled set with 1200 samples. All the results are averaged over 10 different runs.

	Initial	Proposed	MS	RS
# Training set	200	1200		
OA	62.43±2.40	**82.32**±0.41	82.30±0.70	79.75±0.85
AA	66.67±1.63	81.20±0.49	**81.46**±0.97	78.74±1.25
Kappa	0.569±0.027	**0.794**±0.005	0.793±0.008	0.763±0.01

Table 2. The averaged number of classes covered by the newly labeled instances at each iteration and the associated time saving by the proposed incremental strategy. The results are calculated over one of the ten considered runs.

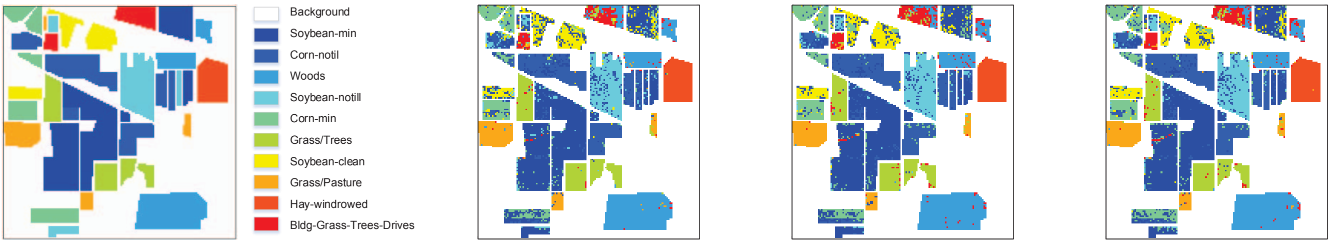
	Proposed	MS	RS
No. Classes covered	4.8±0.9	4.9±1.1	6.0±1.1
Time saving (%)	24.91±9.87	24.25±10.91	14.84±8.75

the experiments: 1) the state-of-the-art MS AL method [6] using OAO multiclass strategy, and 2) the random selection (RS) method which randomly selects the instances from the unlabeled set. In the experiments, The SVM optimization problem was solved using the LIBSVM [9], in which the Gaussian kernel is selected on account of its excellent performance in hyperspectral image classification, and five folds cross-validation was applied to determine the parameters only at the first iteration.

In Fig. 1, we present the classification performance in terms of overall accuracy (OA) during the query steps and the corresponding standard deviation values for each AL method. As can be seen, the AL approaches give faster convergence rate and consistent improvement with respect to the RS method. In particular, the proposed AL method gives competitive performance with respect to the state-of-the-art MS method. Moreover, a lower standard deviation values are obtained by the proposed method, which implies the better stability of the proposed approach compared with the MS and RS methods.

A more detailed accuracy values in terms of OA, average accuracy (AA), and kappa coefficient (Kappa) are shown in Table 1. The classification performances by the SVM classifier trained on the final labeled set with different sample selection methods are provided. For the three methods, a total of 1200 labeled samples were collected respectively by the AL process including the initial 200 labeled sample. We can see from the numerical results in Table 1, the proposed AL method gives the best performance than the MS and RS methods in term of OA and Kappa criterions.

Table 2 gives the evaluation results of the proposed incremental strategy during the AL process. We can see from Ta-



(a) Ground Truth (b) RS: OA= 79.32% (c) MS: OA= 81.98% (d) Proposed: OA= 81.54%

Fig. 2. Classification map achieved by the SVM classifier trained on the final labeled set collected with different AL methods. a) Ground truth image of Indian Pines data. b) Results with random selection (RS) method. c) Results with marginal sampling (MS) method. d) Results with proposed AL approach.

ble 2, as the training processes of a certain number of binary classifier and the corresponding prediction process of the unlabeled instances are avoided, time saving is achieved by the proposed model updating strategy.

Finally, for illustrative purposes, Fig. 2 shows the classification maps on the final labeled set obtained in one of the ten considered runs along with the respective ground-truth map of AVIRIS Indian Pines data. The improvements in classification obtained by the proposed AL approach with respect to the RS method can be graphically observed in Fig. 2.

4. CONCLUSIONS

We have studied the active learning methods for the classification of hyperspectral images. To iteratively enlarge the training set, we proposed an AL technique based on the SVM classifier, in which the OAO strategy was used to construct the multiclass classifier. In the proposed approach, the two mostly voted classes were firstly identified for each instance in the unlabeled set, and then the uncertainness of each instance was measured by the absolute discriminant function value predicted by the binary classifier on the two classes. Finally, the most uncertain instances were selected for manual labeling and then added to the labeled set. In addition, an incremental strategy for model training and prediction was proposed, which could avoid the repeatedly computation during the AL iterations. Experiments were conducted on the AVIRIS Indian Pines data set to evaluate and compare the performance of the proposed AL approach with respect to the state-of-the-art AL method. The results demonstrated the effectiveness of the proposed AL approach and the competitive performance and better stability with respect to the MS AL method. And the results also validated the efficiency of the proposed incremental strategy during the AL process .

As future work, we will test the proposed method on other hyperspectral data sets and compare the proposed AL approach with other typical AL methods in literatures.

5. REFERENCES

[1] Jos M. Bioucas-Dias, Antonio Plaza, Gustavo Camps-Valls, Paul Scheunders, Nasser M. Nasrabadi, and Jocelyn Chanussot, "Hyperspectral remote sensing data analysis and future challenges," *IEEE Geosci. Remote Sens. Mag.*, vol. 1, no. 2, pp. 6–36, Jun. 2013.

[2] G. Hughes, "On the mean accuracy of statistical pattern recognizers," *IEEE Trans. Inf. Theory*, vol. IT-14, no. 1, pp. 55–63, Jan. 1968.

[3] Gustavo Camps-Valls and Lorenzo Bruzzone, "Kernel-based methods for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 6, pp. 1–12, Jun. 2005.

[4] Melba M. Crawford, Devis Tuia, and Hsiuhan Lexie Yang, "Active learning: Any value for classification of remotely sensed data ?," *Proc. IEEE*, vol. 101, no. 3, Mar. 2013.

[5] Devis Tuia, Michele Volpi, Loris Copa, Mikhail Kanevski, and Jordi Muoz-Mar, "A survey of active learning algorithms for supervised remote sensing image classification," *IEEE J. Sel. Topics Signal Process.*, vol. 5, no. 3, pp. 606–617, Jun. 2011.

[6] Edoardo Pasolli, Farid Melgani, Devis Tuia, Fabio Pacific, and William J. Emery, "SVM active learning approach for image classification using spatial information," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 4, pp. 2217–2233, Apr. 2014.

[7] Burr Settles, "Active learning literature survey," Comput. Sci. Tech. Rep. 1648, Univ. WisconsinMadison, 2009.

[8] Chih-Wei Hsu and Chih-Jen Lin, "A comparison of methods for multi-class support vector machines," *IEEE Trans. Neural Netw.*, vol. 13, no. 2, pp. 415–425, Mar. 2002.

[9] C. C. Chang and C. J. Lin, "LIBSVM: a library for support vector machines," 2011, Available: http://www.csie.ntu.edu.tw/ cjlin/libsvm.