ISSN:2320-0790

An International Journal of Advanced Computer Technology

Interactive Domain Adaption for the Classification of Remote Sensing Images Using Active Learning

U.Pushpa Lingam
Department of Computer Science and Engineering,
RVS college of Engineering and Technology,
Coimbatore, India

Abstract: Interactive Domain Adaptation (IDA) technique based on active learning for the classification of remote sensing images. Interactive domain adaptation method is used for adapting the supervised classifier trained on a given remote sensing source image to make it suitable for classifying a different but related target image. The two images can be acquired in different locations and at different times. This method iteratively selects the most informative samples of the target image to be labeled and included in the training set and the source image samples are reweighted or removed from the training set on the basis of their disagreement with the target image classification problem. The consistent information available from the source image can be effectively exploited for the classification of the target image and for guiding the selection of new samples to be labeled, whereas the inconsistent information is automatically detected and removed. This approach significantly reduces the number of new labeled samples to be collected from the target image. Experimental results on both a multispectral very high resolution and a hyper spectral data set confirm the effectiveness of the interactive domain adaptation for theclassification of remote sensing using active learning method.

Keywords: Active learning (AL), domain adaptation (DA), image classification, support vector machine (SVM)

I. INTRODUCTION

The continuously growing availability of remote sensing images gives the opportunity to develop several important applications related to land cover monitoring and mapping. To exploit such an opportunity, it is necessary to develop adequate classification system capable to produce accurate land cover maps atreasonable cost and time. At present, the most common approach to obtain land cover maps is based on supervised learning methods. Supervisedlearning methods require a new set of labeled training samples every time that anew remote sensingimage has to be classified. It leads to high cost for the acquisition of additional reference information. This is due to possible differences in the image acquisition conditions (e.g., illumination and viewing angle), ground conditions (e.g., soil moisture and topography) that may affect the observed spectral signatures of the land cover classes.

In domain adaptation the main goal is to adapt a classifier initially trained allows to exploit with examples coming from a source domain to produce good predictive performances on samples coming from a different but related target domain. Interactive domain adaptation technique for the classification of RS images the consistent information of the source image to classify the target image. This way the amount of target samples to be labeled can be significantly reduced. The proposed method when guided interactive by the classification system by means of active learning (AL) technique that iteratively selects the most informative samples from the target image to be labeled.

The main novel contributions are: 1) the use of a **query+** function that considers both uncertainty and diversity criteria for addressing DA problems; 2) the introduction of a **reweighting** mechanism for source domain samples based on the cosine angle similarity measure in kernel space; 3)the definition of a **query-** function that adaptively selects the inconsistent samples to be discarded.

II. LITERATURE SURVEY

Image processing involves changing the nature of an image in order to either improve its pictorial information for human interpretation and also for autonomous machine perception. Remote sensing images are of great interest in numerous applications. Map drawings, delimitation of parcels, studies on hydrology, forest or agriculture are just a few examples where these images. Many of these applications involve the classificationof pixels in an image into a number of classes.The continuously growing availability of Remote Sensing (RS) images gives us theopportunity to develop several important applications related to land-covermonitoring and mapping. In order to statistically characterize the variation betweenthe source and target domains. In [5],a method for addressing the covariate shift problem byreweighting source-domain samples is proposed. The weightsfor the source-domain samples are obtained by minimizing thediscrepancy between the distributions of the unlabeled samplesin the source and target domain. In [6], an AL techniqueto address data set shift problems in the classification of RSimages under covariate shift is proposed. Finally, an AL methodto address transfer learning problems in the classification of hyperspectraldata has been proposed in [3], employing a samplereweighting method based on the TrAdaBoost algorithm [7].our work isfocused on classify the source and target image compared to the existing one our work has advantage to reduce cost, time, and labeling.

III. ACTIVE LEARNING WITH SVM

Proposed system iteratively adapts the classifier to the target-domain problem. If the two classification problems are highly related, the number of samples of the target image to be annotated can be strongly reduced by exploiting most of the source-domain samples. If the classification problems are less similar, the proposed system will nevertheless allow the classifier to adapt to the target domain. In the proposed approach, the supervised classification is performed using support vector machines (SVMs), which proved very effective in the classification of both multispectral and hyperspectral images.

A SVM is a linear or non-linear classifier, which is a mathematical function that can distinguish two different kinds of objects.

Support vector machine is a supervised learning classification technique to cover land monitoring community.

Analyze data and regression pattern for the classification.

Classify both multispectral image and hyper spectral image.

We adopt a formulation that considers different weights for source-domain instances in the learning phase. More precisely, we solve the following constrained minimization problem:

$$\min_{w,\xi^s,\xi^t,b} \frac{1}{2}\|w\|^2 + C(\sum_{j=1}^{m} \xi_j^t + \sum_{i=1}^{n} \beta_i \xi_i^s)$$

Subject to

$$y_j^t[w.\emptyset(X_j^t) + b] \geq 1 - \xi_i^s \; j = 1, \ldots, m$$
$$y_j^s[w.\emptyset(X_j^s) + b] \geq 1 - \xi_i^s \; i = 1, \ldots, n$$
$$\xi_j^s, \xi_i^t \geq 0$$

Where **w** is a vector orthogonal to the separating hyper plane; bis a bias term such that b/_**w**_ represents the distance of the hyper plane from the origin; Cis the regularization parameter's is the function mapping the data into the feature space; $\xi_i^s$and ξare the slack variables associated with the source- and target domain samples, respectively; mand n are the numbers of target- and source-domain samples at a given iteration, respectively and βiare the weights for source samples.

IV.INTERACTIVE DOMAIN ADAPTATION METHOD

A.Query +

The aim of the query+ function is to select a batch of the most informative samples from a pool U of unlabeled samples, which are taken from the target domain. Once selected, such samples are manually labeled by the user and added to the training set. In our approach, we adopted the batch-mode query function MCLU (i.e., multiclass-level uncertainty)–ECBD (i.e., enhanced clustering-based diversity). Such a technique selects a batch of informative samples from the pool by considering both uncertainty and diversity. The uncertainty criterion is associated to the confidence of the supervised algorithm in correctly classifying the considered samples, whereas the diversity criterion aims at selecting a set of unlabeled samples that are as more diverse (distant to one another) as possible, thus reducing the redundancy among the selected samples

B.Reweight And Remove Source-Domain Samples

The weight for each source-domain sample is computed by considering its similarity to the target-domain samples of the same class according to the mean cosine-angle similarity defined as

$$\beta_i = \frac{1}{m_{y_i^s}} \sum_{j:y_j^t=y_i^s} \frac{k(X_i^s, X_j^t)}{\sqrt{k(X_i^s, X_j^s)k(X_i^t, X_j^t)}}$$

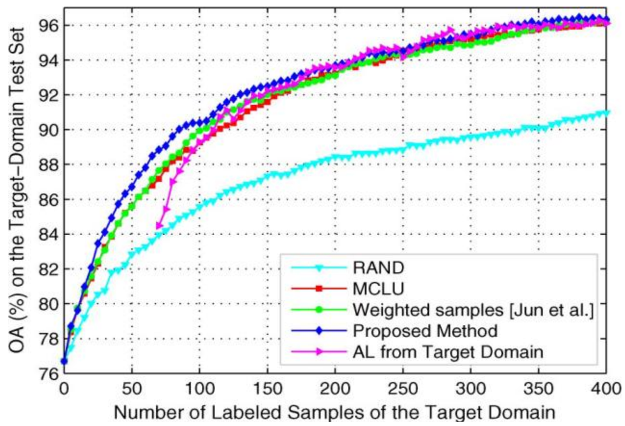
Where $m_{y_i^s}$ is the number of target-domain samples X_j^t associated to the same class $y_j^t = y_i^s$ of the source-domain sample, and $k(\cdot, \cdot)$ is a positive semi definite kernel function. In ourimplementation, we adopted a radial basis function (RBF) kernel (the same as for the SVM classifier). The weights $\beta_i$therefore assume value in the range (0, 1).The rationale of this reweighting procedure is to reduce the weight of source-domain samples that are far apart from the samples of the same class in the target domain.

C. Query-

Query- function removes a fixed amount of source-domain samples at each iteration of the AL process. Here, we adopt instead a simple heuristic to remove the inconsistent source-domain samples from that training set that does not require fixing a priori the amount of samples to be removed at each iteration. In the proposed methodology, we remove at each iteration the source domain samples that are misclassified by the SVM classifier. This is done by setting $\beta_i = 0$ in correspondence with the misclassified source-domain samples. Summarizing, at each iteration of the AL process, the new labeled samples selected by the query+ function are included in the training set, the weights $\beta_i$of source-domain samples are recomputed considering the reweighting procedure and the query function, and the SVM algorithm is iterated again and again to achieve better performance.

V.EXPERIMENTAL RESULTS

The obtained results show that the proposed technique leads to significantly higher accuracy values than standard methods, confirming its effectiveness in exploiting the consistent information of the source image and in removing the inconsistent one for classifying the target image. The proposed method using the reweighting heuristic presented in at different iterations.The experiments are carried out on both a multispectral very high resolution (VHR) and a hyperspectral data set.



Overall Accuracy

This shows the overall accuracy The different curves correspond to: 1) random selection; 2) the MCLU AL method; 3) a method that combines MCLU and the reweighting procedure presented in [3]; 4) the proposed method; and 5) the MCLU–ECBD AL method applied directly to the target domain.

VI.CONCLUSION

IDA method for the classification of RS images has been proposed. The proposed method allows the user to effectively exploit the consistent information of a source image for the classification of a different but related target image. This can result in a significant reduction of the number of new target-domain samples to be labeled, thus reducing the cost associated with the classification of the target image. In operative scenarios, when the budget for acquiring new labeled samples is limited, the user may decide to stop the IDA procedure at early iterations as soon as the desired level of accuracy is reached. The experimental results obtained in the classification of both a multispectral VHR and hyperspectral images confirm the effectiveness of the proposed technique.

VII. REFERENCES

[1] P. Mitra, B. U. Shankar, and S. K. Pal, “Segmentation of multispectral remote sensing images using active support vector machines,” *PatternRecognit. Lett.* vol. 25, no. 9, pp. 1067–1074, Jul. 2004.

[2] S. Rajan, J. Ghosh, and M. Crawford, "An active learning approach tohyperspectral data classification," *IEEE Trans. Geosci. Remote Sens.*,vol. 46, no. 4, pp. 1231–1242, Apr. 2008.

[3] G. Jun and J. Ghosh, "An efficient active learning algorithm with knowledgetransfer for hyperspectral data analysis," in *Proc. IEEE IGARSS*,Jul. 2008, vol. 1, pp. I-52–I-55.

[4] B. Demir, C. Persello, and L. Bruzzone, "Batch-mode active-learningmethods for the interactive classification of remote sensing images," *IEEETrans. Geosci. Remote Sens.*, vol. 49, no. 3, pp. 1014–1031, Mar. 2011.

[5] J. Huang, A. Gretton, B. Schölkopf, A. J. Smola, and K. M. Borgwardt,"Correcting sample selection bias by unlabeled data," in *Proc. Advancesin Neural Information Processing Systems*. Cambridge,MA: MIT Press, 2007.

[6] D. Tuia, E. Pasolli, and W. J. Emery, "Using active learning to adaptremote sensing image classifiers," *Remote Sens. Environ.*, vol. 115, no. 9, pp. 2232–2242, Sep. 2011.

[7] W. Dai, Q. Yang, G. Xue, and Y. Yu, "Boosting for transfer learning," in*Proc. Int. Conf. Mach. Learn.*, 2007, pp. 193–200.

[8] C. Persello and L. Bruzzone, "Active learning for domain adaptation in thesupervised classification of remote sensing images," *IEEE Trans. Geosci.Remote Sens.*, vol. 50, no. 11, pp. 4468–4483, Nov. 2012.

[9] B. Schölkopf and A. J. Smola, *Learning with Kernels: Support VectorMachines, Regularization, Optimization, and Beyond*. Cambridge, MA: MIT Press, 2001.

[10] F. Melgani and L. Bruzzone, "Classification of hyperspectral remote sensingimages with support vector machines," *IEEE Trans. Geosci. RemoteSens.*, vol. 42, no. 8, pp. 1778–1790, Aug. 2004.