DOMAIN SEPARATION FOR EFFICIENT ADAPTIVE ACTIVE LEARNING

Giona Matasci[1], *Devis Tuia*[2], *Mikhail Kanevski*[1]

[1] Institute of Geomatics and Analysis of Risk, University of Lausanne, Lausanne, Switzerland.
[2] Image Processing Laboratory (IPL), Universitat de València, València, Spain.
Email: giona.matasci@unil.ch

ABSTRACT

This paper proposes a procedure aimed at efficiently adapting a classifier trained on a source image to a similar target image. The adaptation is carried out through active queries in the target domain following a strategy particularly designed for the case where class distributions have shifted between the two images. We first suggest a pre-selection of candidate pixels issued from the target image by keeping only those samples appearing to be lying in a region of the input space not yet covered by the existing ground truth (source domain pixels). Then, exploiting a classifier integrating instance weights, active queries are performed on the target image. As the inclusion to the training set of the samples progresses, the weights associated with the training pixels are updated using different criteria according to their origin (source or target domain). Experiments on a pair of QuickBird images of urban scenes prove the validity of the proposed approach if compared to existing benchmark methods.

Index Terms— image classification, domain adaptation, domain separation, active learning, TrAdaBoost

1. INTRODUCTION

It is well known that the success of remote sensing image classification algorithms strongly depends on the quality of the training data used. However, the ground truth collection process is not a trivial task: the procedure requires either expensive terrain campaigns or time-consuming photo-interpretation analyses.

Active learning (AL) techniques have recently been applied in this context to guide the user in the optimal selection of the pixels to be sampled [1]. The purpose of such techniques is to build a training set yielding a classification model able to efficiently discriminate the land cover classes using the lowest possible number of labeled pixels. Therefore, AL methods are requested to improve a passive selection of the samples to label by providing the user with the most informative pixels, especially during the first iterations of the algorithm.

Recently, it has been shown that the labeling effort could be further reduced by making use of already collected ground truth associated with images acquired by the same sensor in a region with comparable characteristics. In fact, it would be beneficial for the user to apply routines to adapt a classifier designed to model a given image, the source domain, to another image sharing the same spectral channels and classes to be described, but governed by slightly different probability distributions. We will refer to this second image as the target domain. The study of adaptation algorithms is referred to as *domain adaptation* (DA) [2] and has been recently tackled in the remote sensing community [3, 4, 5, 6].

These recent works tried to estimate the class distribution in the target domain by adapting the source classifier via unsupervised or semisupervised techniques. Particularly, these methods assumed that the labeled examples from the target image, when available, were passively obtained at once. However, it has been shown in the context of hyperspectral image classification that DA can also be achieved by actively querying the target samples to label, to be successively integrated in the knowledge transfer process [7].

This paper also proposes a combination of the two frameworks, taking as starting point the previously mentioned work by Jun *et al.* and the one presented in [8] for the correction of sample selection bias in remote sensing image classification.

We extend these DA strategies by integrating the approach proposed in [9]. Rai *et al.* suggest a preprocessing step highlighting the interesting regions of the target domain in order to reduce the size of the set of candidate pixels. An AL approach based on the same concept has recently been proposed within the remote sensing community in [10]. In the context of SVM classification, the authors advise to identify significant uncertain regions of the input space by training a model to discriminate Support Vectors from the rest of the samples.

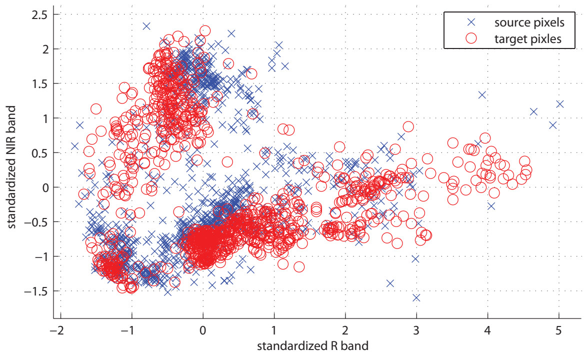
Then, after the preprocessing stage, we propose to sample the most informative pixels with active queries from the target image while adapting the obtained classifier using a transfer learning strategy to leverage the original source data [11]. Such an approach is intended to boost the performance of traditional AL techniques when asked to intelligently suggest a sampling scheme in a target image whose class distributions have shifted.

2. DOMAIN SEPARATION AND ADAPTIVE ACTIVE LEARNING

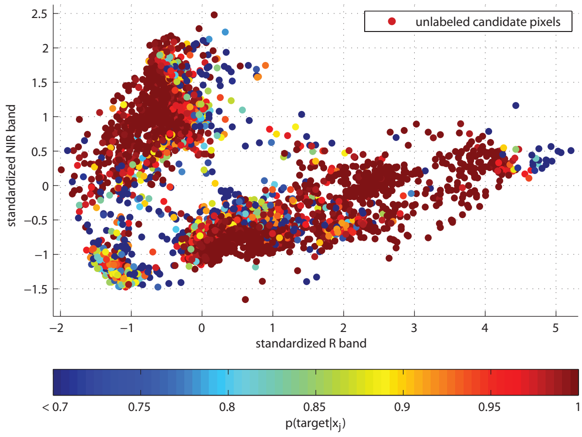
The adaptation of the classifier is achieved by performing active queries in the target domain.

In this context, we first apply the *domain separation* (DS) principle (*domain separator hypothesis*) suggested by Rai *et al.* [9]. This technique consists in avoiding the acquisition of labels for samples lying in a region of the input space already covered by the source dataset. With this purpose, we perform a binary classification aiming at the separation of the source and target examples. Indeed, if the two domains overlap, some of the candidates (belonging to the target domain)

This work has been supported by the Swiss National Science Foundation (grants no. 200021-126505 and PBLAP2-127713/1).



(a) The region occupied by the 1000 source pixels of the training set (blue crosses) is compared to the region spanned by 1000 target pixels randomly selected from the unlabeled set U (red circles).



(b) Posterior probabilities of target domain membership for the set of unlabeled candidates (for readability sake only 1/5 of these 22'723 samples is plotted). Low target domain probabilities (blue tones) suggest pixels whose removal would benefit the active selection of target samples by constraining the region searched.

Fig. 1. Example of scatterplots of red versus near infrared QuickBird bands (using data of the 1st experiment) to depict the DS concept.

will be found on the source side. In this case, the samples are not worth the labeling effort since they will bring in redundant information. This principle is illustrated by Fig. 1 making use of the real QuickBird data analyzed in this study. The dataset shift occurred between the two acquisitions is relatively evident: it allows us to classify one domain against the other in order to obtain target domain probabilities for the set of unlabeled candidates. Throughout this work we consider SVM posterior probabilities estimated with the Platt's method. One can remark the correspondence between the distribution of these probabilities shown in Fig. 1(b) and class boundaries of Fig. 1(a).

Afterwards, we apply the *breaking ties* algorithm (BT) [12], which is a state-of-the-art active learner relying on the uncertainty of classification posterior probabilities. We then combine this AL strategy with a transfer learning technique directly inspired by the *TrAdaBoost* procedure presented in [11]. For this purpose, we exploit a SVM integrating instance weights in the training phase (see `http://www.csie.ntu.edu.tw` for more information as well as [13] for theoretical foundations).

Concretely, the proposed strategy first comprises the DS preprocessing step:

- We remove from the unlabeled set of candidates $U = \{\mathbf{x}_j\}_{j=1}^{u}$ the examples $\mathbf{x}_j$ whose probability to belong to the target domain $p(\text{target}|\mathbf{x}_j)$ is lower than a certain threshold p_T. These $p(\text{target}|\mathbf{x}_j)$ values are the probabilistic outputs of a SVM classifier trained with samples randomly selected from the target image (labeled as $+1$) and the source image (labeled as -1). Therefore, only candidate pixels providing useful new insights about the class distribution in the target domain will be queried in the AL procedure.

Then, two nested loops are run to achieve DA:

- In the outer AL loop of the algorithm, during each iteration, the available labeled training set T is divided into two subsets: a source set of n samples $T_S = \{(\mathbf{x}_i, y_i)\}_{i=1}^{n}$ and a target set of m samples $T_T = \{(\mathbf{x}_i, y_i)\}_{i=n+1}^{n+m}$. Initially $T_T = \{\}$, then, at each iteration of the AL process, this set of target training points is extended with the q most interesting candidates $\mathbf{x}_j$ (after determining the corresponding label y_j) identified in the reduced target unlabeled set U using BT. This heuristic selects the best points $\hat{\mathbf{x}}^{BT}$ according to the following ranking criterion:

$$\hat{\mathbf{x}}^{BT} = \arg\min_{\mathbf{x}_j \in U} \Big(\max_{cl \in \Omega} p(y_j^* = cl|\mathbf{x}_j) - \max_{cl \in \Omega \backslash cl^+} p(y_j^* = cl|\mathbf{x}_j) \Big), \quad (1)$$

  where $cl^+ = \arg\max_{cl \in \Omega} \left(p(y_j^* = cl|\mathbf{x}_j) \right)$ is the class with the highest probability for pixel $\mathbf{x}_j$ and $\Omega = \{cl_1, \ldots, cl_C\}$ is the set of C classes. A SVM model able to weight the training samples, with coefficients $\{w_i\}_{i=1}^{n+m}$, is used to compute class probabilities $p(y_j^* = cl|\mathbf{x}_j)$.

- At each round of the AL cycle, TrAdaBoost is run to iteratively reweight the training samples in T. Starting with uniform weights $w_i^t = \frac{1}{(n+m)}, \forall\, i$ at the boosting iteration $t = 0$, we update the weights of the instances for the subsequent boosting iteration in two distinct ways, according to their origin, as proposed in [11]. The principle consists in re-predicting labels y_i for T using this same training set T with weighted samples. To update the current weights and obtain w_i^{t+1}, training errors are taken into account: if the sample is correctly classified, $w_i^{t+1} = w_i^t$. If the sample is misclassified, two options are possible: if $\mathbf{x}_i \in T_S$, the weight is decreased by a constant factor. On the contrary, if $\mathbf{x}_i \in T_T$, the weight is increased by a factor inversely proportional to the training error computed on T_T. This aims at reducing the influence of misleading source examples, supposed to be the most dissimilar to the target instances we are interested in. Conversely, the augmentation of the influence of misclassified target samples in the decision function translates the need to focus on the regions of the target domain in which the classification is hard to solve, thus being beneficial for the AL loop. The process is run until a convergence criterion is reached. Once updated, the weights are used to retrain the instance weighting SVM yielding the class probabilities used by BT to perform the active selection on the pool of candidates U. This iterative approach differs from the one adopted in [7] since, in the latter, the weights to be used in the next AL iteration are not the result of a repeated updating process, the TrAdaBoost loop, but are computed in a single stage.

By applying this inductive transfer learning approach we let the SVM model gradually adapt to the target domain. First, the DS step allows us to identify the most helpful samples from the set of candidates, ensuring thus that the active queries will not be performed on redundant pixels. Afterwards, the different weighting of the instances leads to a decision function becoming more and more suited to discriminate the classes in the target domain. Hence, this procedure affects both the performance of the classification on test data (belonging exclusively to the target image) and the correctness of class membership probabilities computed for the unlabeled samples in U. The latter effect induces a selection criterion better suited to identify candidates lying in uncertain regions of the target input space.

3. DATA AND EXPERIMENTAL SETUP

We tested the proposed methodology on a dataset consisting of two VHR QuickBird images (acquired in 2002 and 2006) of the city of Zurich, Switzerland, representing two different neighborhoods. The one considered as the target was taken in August while the one considered as the source was acquired in October. Differences in illumination conditions and seasonal effects on the vegetation affect the class-conditional distributions.

The pansharpened images have a spatial resolution of 0.6 m and present 4 bands covering the region of the spectrum from 450 to 900 μm. The histograms have first been matched and, subsequently, textural (3×3 data range, mean, homogeneity and entropy) and morphological (5×5 opening and closing, 7×7 and 9×9 opening and closing by reconstruction) features have been extracted from the panchromatic band. The total number of considered features was 15 (4 MS bands, 1 PAN band, 4 textural f., 6 morphological f.).

By visual inspection, we identified and labeled pixels from 4 ground cover classes characterizing both images: “Buildings”, “Roads”, “Grass” and “Vegetation”. Additionally, a class “Shadows” has been added. The initial training set, exclusively built from the source image, was composed of 15’934 pixels while the unlabeled set of candidates was extracted from the target image and included 22’723 pixels. The generalization ability of the different techniques in the target domain was assessed on 26’797 test samples issued from spatially separated regions of the target image.

For the proposed method (`DS_AdaptiveAL_BT`), the classification of the source domain against the target domain was performed with equal class sizes and candidate pixels with a probability to belong to the target domain lower than $p_T = 0.8$ have been discarded. Always concerning this technique, at each AL iteration, the weights of the samples in the training set were updated after 8 iterations of TrAdaBoost (stabilized w_i values). In this sub-routine, the re-prediction of labels for the training samples was implemented through a 20-fold cross-validation to avoid overfitting. We compared the proposed DS strategy with the method not using any domain separator hypothesis originally proposed in [7] (`JunAdaptiveAL_BT`), the standard BT without instance reweighting (`AL_BT`) and a random selection of the samples to label (`RandomS`). Moreover, to set reference accuracies for the considered target image, classifiers trained exclusively on target and source datasets have been tested (`Target` and `Source` methods).

The experiments were conducted with 10 different initializations of the training sets, each one with 1000 randomly selected pixels. A SVM with Gaussian kernel has been utilized as supervised learner and a 10-fold cross-validation has been performed to find the optimal initial C and σ parameters. For all the AL methods, $q = 10$ target samples per iteration over 34 iterations were added to augment the initial source training set.

The algorithms were implemented in MATLAB using LIBSVM as library both for the standard SVM and instance weighting SVM [14]. The computation of class probabilities to be used by BT is described in the same paper.

4. RESULTS

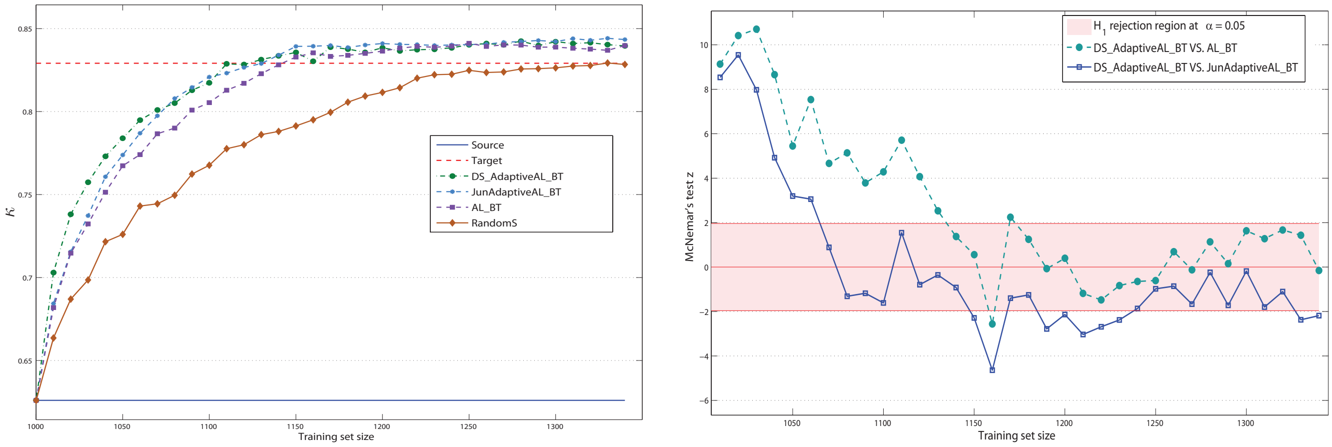
Figure 2 presents a summary of the results obtained for this task of DA through active learning. A comparison of the tested strategies is supplied by means of the learning curves of Fig. 2(a).

First, one can notice the relatively bad performance achieved by applying the source model (`Source`) without any adjustments: average estimated Kappa statistic κ equals to 0.626. Then, looking at the baseline for AL, the method consisting in randomly sampling the pool of unlabeled pixels (`RandomS`) shows a slow convergence. Only after having added 330 samples from the target domain to T, we are able to reach the performance of a SVM trained with pixels from the target image only (`Target` reference accuracy of $\kappa = 0.829$).

The proposed combined methodology integrating the DS concept (`DS_AdaptiveAL_BT`) clearly outperforms this baseline by increasing the classification accuracy since the very beginning of the AL iterations. The comparison with the non-adaptive AL heuristic of BT (`AL_BT`) shows a steeper ascent of the curve for the proposed adaptive method, able to reach the upper reference accuracy by adding a smaller number (110 VS. 140) of pixels to the base training set.

Looking at the general behavior of the learning rates for the two adaptive AL methods, we observe the `DS_AdaptiveAL_BT` curve as always evolving in between 0.01 and 0.02 κ higher than the `JunAdaptiveAL_BT` curve until the 6^{th} AL iteration.

More reliable conclusions can be drawn from Fig. 2(b). The plot illustrates the assessment of the statistical significance of models’ performances using the two-tailed McNemar’s test z value. Taking the standard α level of 5%, one can remark the significant superiority ($z > 1.96$) of the proposed DS AL technique over the two other methods during the first key iterations. Such a difference is lasting until the inclusion in T of 60 target pixels when compared to `JunAdaptiveAL_BT` while being noticeable until convergence (until obtaining a training set T of 1130 pixels) when compared to the `AL_BT` method.



(a) Average learning curves over 10 runs. `Source` (solid blue line) = model built using pixels of the source domain only, `Target` (dashed red line) = model built using pixels of the target domain only, `DS_AdaptiveAL_BT` (dashed green line) = proposed DS adaptive AL method, `JunAdaptiveAL_BT` (dashed light blue line) = adaptive AL method proposed in [7], `AL_BT` (dashed purple line) = breaking ties method [12], `RandomS` (solid brown line) = random sampling method.

(b) Evolution of the average McNemar's test z value when assessing the difference between the proposed method and the techniques `JunAdaptiveAL_BT` (solid blue line) and `AL_BT` (dashed green line). A positive sign of the decision value means a better performance by `DS_AdaptiveAL_BT` while values of $|z| > 1.96$ (outside the shaded red region) indicate a statistically significant difference in accuracy at $\alpha = 5\%$.

Fig. 2. Comparison of methods on the Zurich target image (26'797 pixels in the test set).

5. CONCLUSIONS

We presented an efficient approach to boost the performance of active learning methods when applied in the framework of DA. We introduced a rather intuitive technique aimed at ignoring redundant samples in AL strategies when used in a knowledge transfer framework. Indeed, the methodology using a reduced set of suitable candidates and properly adapting training samples weights has proved to be able to perfect the ranking criterion for the selection of the most informative target pixels to be labeled. The user facing a classification problem involving a target image where ground truth collection requires large efforts will be efficiently guided in the sampling process, if already classified source images with similar properties are available.

6. REFERENCES

[1] D. Tuia, M. Volpi, L. Copa, M. Kanevski, and J. Muñoz-Marí, "A Survey of Active Learning Algorithms for Remote Sensing Image Classification," *IEEE J. Sel. Topics Sign. Proc.*, vol. 5, no. 3, pp. 1–12, 2011.

[2] S. J. Pan and Q. Yang, "A Survey on Transfer Learning," *IEEE Trans. Knowl. Data Eng.*, vol. 22, pp. 1345–1359, 2010.

[3] L. Bruzzone and D. F. Prieto, "Unsupervised retraining of a maximum likelihood classifier for the analysis of multitemporal remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 39, no. 2, pp. 456–460, 2001.

[4] L. Bruzzone and M. Marconcini, "Toward the automatic updating of land-cover maps by a domain-adaptation SVM classifier and a circular validation strategy," *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 4, pp. 1108–1122, 2009.

[5] L. Gomez-Chova, G. Camps-Valls, L. Bruzzone, and J. Calpe-Maravilla, "Mean Map Kernel Methods for Semisupervised Cloud Classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 1, pp. 207–220, 2010.

[6] S. Rajan, J. Ghosh, and M.M. Crawford, "Exploiting Class Hierarchies for Knowledge Transfer in Hyperspectral Data," *IEEE Trans. Geosci. Remote Sens.*, vol. 44, no. 11, pp. 3408–3417, 2006.

[7] G. Jun and J. Ghosh, "An efficient active learning algorithm with knowledge transfer for hyperspectral data analysis," in *IEEE International Geoscience and Remote Sensing Symposium, 2008. IGARSS 2008.*, 2008, vol. 1, pp. I–52 –I–55.

[8] D. Tuia, E. Pasolli, and W. J. Emery, "Using active learning to adapt remote sensing image classifiers," *Remote Sensing of Environment*, vol. 115, 2011.

[9] P. Rai, A. Saha, H. Daumé III, and S. Venkatasubramanian, "Domain Adaptation meets Active Learning," in *NAACL-HLT 2010 Active Learning for NLP Workshop*, Los Angeles, USA, 2010.

[10] E. Pasolli, F. Melgani, and Y. Bazi, "Support vector machine active learning through significance space construction," *IEEE Geoscience and Remote Sensing Letters*, vol. PP, no. 99, pp. 431–435, 2010.

[11] W. Dai, Q. Yang, G.-R. Xue, and Y. Yu, "Boosting for transfer learning," in *Proceedings of the 24th International Conference on Machine learning*, New York, NY, USA, 2007, ICML '07, pp. 193–200, ACM.

[12] T. Luo, K. Kramer, D. B. Goldgof, L. O. Hall, S. Samson, A. Remsen, and T. Hopkins, "Active Learning to Recognize Multiple Types of Plankton," *Journal of Machine Learning Research*, vol. 6, pp. 589–613, 2005.

[13] M.W. Chang, C.J. Lin, and R.C. Weng, "Analysis of switching dynamics with competing support vector machines," *IEEE Trans. Neural Networks*, vol. 15, pp. 720–727, 2004.

[14] C.-C. Chang and C.-J. Lin, *LIBSVM: a library for support vector machines*, 2001, Software available at `http://www.csie.ntu.edu.tw/~cjlin/libsvm`.