MULTI-VIEW ADAPTIVE DISAGREEMENT BASED ACTIVE LEARNING FOR HYPERSPECTRAL IMAGE CLASSIFICATION

Wei Di, Student Member, IEEE and Melba M. Crawford, Fellow, IEEE

School of Civil Engineering
Purdue University, West Lafayette, IN 47907, USA
{wdi, mcrawford}@purdue.edu

1. INTRODUCTION

Obtaining labeled data for supervised classification of remotely sensed imagery is expensive and time consuming. Further, manual selection of the training set is often subjective and tends to induce redundancy into the supervised classifier, thus considerably slowing the training phase. Active learning (AL) integrates data acquisition with the classifier design by ranking the unlabeled data to provide advice for the next query which has the highest training utility [1][2][3]. Thus, it explores the maximum potential of the learner toward both the labeled and unlabeled data, and the training set can be maintained as small as possible and focused on the most representative samples for the entire data space. This potentially leads to greater exploitation of the information in the data, while significantly reducing the cost of data collection.

Although active learning has been widely studied in document retrieval and natural language learning [4][5], related research has been extremely limited in remote sensing (see [2] for a brief literature summary). Rajan *et al*. [1] and Jun and Ghosh [3] investigated AL using a hierarchical classification framework. To avoid the problem of high dimensionality, Rajan *et al*. [1] first applied dimension reduction to the data, which may result in information loss for classification. Tuia *et al*. [2] proposed two methods. The first is an extension of the SVM margin sampling that further incorporates data distribution, while the second is an entropy based extension of query-by-bagging algorithm. However, because the committee is generated by bagging, the candidates selected may not be most relevant for decreasing the classification error and be representative of the true data space.

The hundreds of narrow spectral bands in hyperspectral data provide a natural way to construct a diverse and independent committee that can better describe the entire data space while exploiting the "value of disagreement" in a multi-view framework. Generating views by segmenting the data into several disjoint contiguous sub-band sets avoids the risk of biased sampling by bootstrap, as well as the small sample problem associated with high dimensionality. Moreover, multi-view based learning typically converges quickly to the target concept which can greatly reduce the number of required labeled samples [4]. Also there is no assumption about the properties of the base learner, which allows selection of the best base learner for different types of data.

Co-testing [4], the first multi-view active learning method, queries all the samples with at least two-view disagreement, which results in a much looser confliction level when the number of views is larger. Further, it lacks a delicate measurement for the strength and weakness of each view, which is not ideal for hyperspectral image data since different spectral ranges (views) may have different discriminative ability.

An active learning method based on a multi-view adaptive weighted disagreement measure (AMD-WVE) is proposed to attack the two key problems in multi-view based active learning: 1) view generation by utilizing the intrinsic spectral correlation of the hyperspectral image; 2) contention pool generation and pruning by an adaptively quantified disagreement measure coupled with evaluation of the discriminative capability of each view towards each class.

2. MULTIVIEW GENERATION

Denote each hyperspectral pixel vector as $\mathbf{x}$ and the label set $\{\omega_1, \omega_2, \ldots, \omega_c\}$ with n_c classes. The purpose is to find the correct label: $\hat{y} = \arg\max_{y'} f(\mathbf{x}, y', \{D_L, D_U\})$, where D_L is the labeled data pool with n_L samples, and D_U the unlabeled data pool with n_U samples. The available attributes are decomposed into disjoint n_v sets as different views, and an instance is viewed as $(\mathbf{x}_i^1, \mathbf{x}_i^2, \ldots, \mathbf{x}_i^{n_v})$. Two basic requirements [4] for views: compatibility and independence, are loosely obtained by segmenting the data into several disjoint contiguous sub-band sets along the spectral dimension according to the band correlation index. Each sub-band set has lower correlation with other sub-band sets. Since views are weakly dependent, the ratio of contention points to unlabeled samples wii represents an upper-bound of the learning error for pair-wise views [6]. Different spectral ranges contain different information, thus the diversity of "views" is satisfied. The use of both the labeled and unlabeled data to obtain the correlation coefficients further improves the generalization ability. The lower dimensionality of each view also mitigates the problem of high dimensionality relative to the small number of training samples.

3. PROPOSED METHODOLOGY

The estimate of the label of a sample $\mathbf{x}_i \in D_U$ from view v is obtained by $\hat{y}_i = \arg\max_{y'} f(\mathbf{x}_i, y')$. The disagreement level is defined as the number of different estimates: $DL_i = \left|\bigcup_{v=1}^{n_v} \left\{\hat{y}_i^v\right\}\right|$. The Adaptive Maximum Disagreement Contention pool C_{AMD} is constructed by only selecting unlabeled samples with the maximum disagreement level:

$$C_{AMD} = \left\{x_i : \max_{x_i} \left|\bigcup_{v=1}^{n_v} \left\{\hat{y}_i^v\right\}\right|\right\} \tag{1}$$

Samples in C_{AMD} represent the maximum disagreement thus contain the most uncertainty information. Querying samples from it will bootstrap views both to best learn the training set, and to "agree" with each other on the extra unlabeled data. The more views that agree, the smaller the v-intersection of the hypotheses generated from different views, by which the upper bound of the generalization error can be reduced. As learning progresses, the disagreement level will decrease. To avoid the inflating of the contention pool, we use the weighted voting entropy (WVE) to quantitatively measure the "uncertainty" of the given votes by each view, which also incorporates the differences in discrimination ability of each view with respect to each class.

First initialize the $n_v \times n_c$ weighting matrix with each entry $\mathbf{W}_{vc}\left(v,c\right)=1$ and let $\varpi^{\tau} = \sum_{v=1}^{n_v}\sum_{c=1}^{n_c}\mathbf{W}_{vc}^{\tau-1}\left(v,c\right)$, $\delta_{\omega_c}\left(\hat{y}\right)=\begin{cases}1, \hat{y}=\omega_c \\ 0, \hat{y}\neq\omega_c\end{cases}$. $\mathbf{W}_{vc}^{\tau-1}\left(v,c\right)$ is the weighting matrix of last query, and $wve_{\omega_c}^{\tau}\left(\mathbf{z}_i\right)=\sum_{v=1}^{n_v}\mathbf{W}_{vc}^{\tau-1}\left(v,c\right)\times\delta^{\omega_c}\left(f^v\left(\mathbf{z}_i\right)\right)$, the weighted voting entropy at the τth query for $\mathbf{z}_i \in C_{AMD}$ is:

$$WVE^{\tau}(\mathbf{z}_i) = -\frac{1}{\log\varpi^{\tau}}\sum_{c=1}^{n_c}\frac{wve_{\omega c}^{\tau}(\mathbf{z}_i)}{\varpi^{\tau}}\log\left(\frac{wve_{\omega c}^{\tau}(\mathbf{z}_i)}{\varpi^{\tau}}\right) \tag{2}$$

Only samples with higher entropy values will be further selected into the second stage contention pool C_{WVE} :

$$C_{WVE}^{\tau} = \left\{\mathbf{q}_i : WVE^{\tau}(\mathbf{q}_i) \geq WVE^{\tau}(\tilde{\mathbf{q}}_i)\right\} \tag{3}$$

where $\mathbf{q}_i \in C_{WVE}^{\tau}$, $\tilde{\mathbf{q}}_i \in \bar{C}_{WVE}^{\tau}$, and $C_{WVE}^{\tau} \cup \bar{C}_{WVE}^{\tau} = C_{AMD}^{\tau}$. The desired ratio of the size of C_{WVE}^{τ} to C_{AMD}^{τ} can be set by a threshold. Then query the next n_Q samples by random sampling (denoted as ξ_R) from C_{WVE}^{τ} :

$$\left\{\mathbf{q}_1, \mathbf{q}_2, \ldots, \mathbf{q}_{N_Q}\right\}^{\tau} = \xi_R\left(C_{WVE}^{\tau}\right) \tag{4}$$

Then $\mathbf{W}_{vc}$ is updated by the (0-1) loss rule and the feedback information from the queried samples:

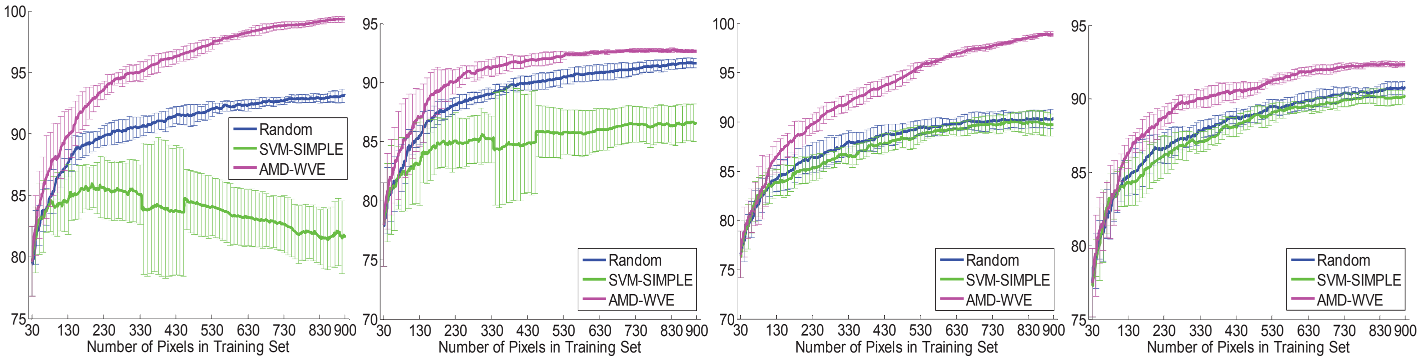
$$\mathbf{W}_{vc}^{\tau}\left(v,c\right) = \mathbf{W}_{vc}^{\tau-1}\left(v,c\right) + \sum_{i_Q=1}^{n_Q}\delta^{y_{i_Q}^{\tau}}\left(\mathbf{Y}^{\tau}\left(v,i_Q\right)\right) \tag{5}$$

where $1 \leq i_Q \leq n_Q$, each column of $\mathbf{W}_{vc}^{\tau}$ is further scaled to sum to one, and $\mathbf{Y}^{\tau}$ is a $n_v \times n_Q$ matrix that contains the labels estimated by each view before the τth query corresponding to the queried n_Q samples. Each entry is:

$$\mathbf{Y}^{\tau}\left(v,i_Q\right) = \hat{y}^v = \arg\max_{y'} f^v\left(\mathbf{q}_{i_Q}^v, y', \{D_L^{v,\tau}, D_U^{v,\tau}\}\right) \tag{6}$$

4. EXPERIMENT

Two AVIRIS data sets (KSCI and KSCII) at 18-m spatial resolution with 10 classes are used here. Five views were generated. Algorithms ran for 10 x cross-validation and 870 epochs each time, adding the pixel with the highest WVE value to the labeled training set at each query. Classification result was obtained by training the SVM base learner using the whole spectral range. The obtained weighting matrix shows that the discriminative ability of each view differs and is class dependent. Also, as learning progresses, the overall classification accuracy of each view for the unlabeled data clearly tends to agree with each other, and the maximum WVE value decreases, indicating that the degree of confusion of the unlabeled data for the learner committee decreases. Our proposed algorithm superior outperforms the random selection and SVM_{SIMPLE} both on the unlabeled data and the unseen data (Fig. 1).



(a) (b) (c) (d)

Fig.1. Classification accuracy for unlabeled data (a) KSCI (c) KSCII and for unseen data (b) KSCI (d) KSCII.

5. CONLUSTIONS

A multi-view based active learning method AMD-WVE is proposed which utilizes the intrinsic multi-view character of the hyperspectral data and adaptively and quantitatively measuring the disagreement level. Further it incorporates the different discriminative ability of different spectral ranges towards each class into the learning process. The proposed method shows substantially superior results compared to random selection and SVM_{SIMPLE} on both the unlabeled and unseen data from two AVIRIS hyperspectral image data sets with 10 classes.

6. REFERENCES

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

[2]D. Tuia, F. Ratle, F. Pacifici, M. F. Kanevski, and W. J. Emery, “Active Learning Methods for Remote Sensing Image Classification”, *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 7, pp. 2218–2232, July. 2009.

[3] G. Jun, J. Ghosh, “An Efficient Active Learning Algorithm with Knowledge Transfer for Hyperspectral Data Analysis”, *IEEE International Geosci. and Remote Sens.Symposium,* vol. 1, pp:52-55, July 2008.

[4] 5I. Muslea, S. Minton, and C. A. Knoblock. “Active Learning with Multiple Views”. *Journal of Artifical Intelligence Research*, vol. 27, pp 203-233, 2006

[5] S. Tong , D. Koller, “Support vector machine active learning with applications to text classification,” *The Journal of Machine Learning Research archive*, vol.2, pp.45-66, March 2002.

[6] S. Abney, “Bootstrapping,” In *Proceedings of 40th Meeting of the Association for Computational Linguistics*, pp. 360–367, 2002.