Title: A Batch Mode Active Learning Technique Based on Multiple Uncertainty for SVM Classifier

Author(s): Swarnajyoti Patra and Lorenzo Bruzzone

A Batch Mode Active Learning Technique Based on Multiple Uncertainty for SVM Classifier

Swarnajyoti Patra and Lorenzo Bruzzone, *Fellow, IEEE*

Department of Information Engineering and Computer Science,

University of Trento, Via Sommarive 14, I-38123 Trento, Italy

Email: patra@disi.unitn.it, lorenzo.bruzzone@ing.unitn.it

Abstract

In this paper we present a novel batch mode active learning technique for solving multiclass classification problems by using the support vector machine (SVM) classifier with the one-against-all (OAA) architecture. The uncertainty of each unlabeled sample is measured by defining a criterion which not only considers the smallest distance to the decision hyperplanes, but also takes into account the distances to other hyperplanes if the sample is within the margin of their decision boundaries. To select batch of most uncertain samples from all over the decision region, the uncertain regions of the classifiers are partitioned into multiple parts depending on the number of geometrical margins of binary classifiers passing on them. Then a balanced number of most uncertain samples are selected from each part. To minimize the redundancy and keep the diversity among these samples, the kernel k-means clustering algorithm is applied to the set of uncertain samples and the representative sample (medoid) from each cluster is selected for labeling. The effectiveness of the proposed method is evaluated by comparing it with others batch mode active learning techniques existing in the literature. Experimental results on two different remote sensing data sets confirmed the effectiveness of the proposed technique.

Index Terms

Active learning, query function, SVM, hyperspectral imagery, multispectral imagery, remote sensing.

I. INTRODUCTION

In the literature, many supervised methods have been proposed for classification of remotely sensed data. The classification results obtained by these methods rely on the quality of the labeled samples used for learning. To obtain proper labeled samples is usually expensive and time consuming. Moreover, the manual selection of the training samples often introduces redundancy into the training set of the classifier, thus slowing the training phase considerably without adding relevant information. In order to reduce the cost of labeling and optimize the performance of the classifier, the training set should be as small as possible by avoiding redundant samples and including only most informative patterns (which have highest training utility). Active learning is an approach that addresses this problem. The learning process repeatedly queries unlabeled samples to select the most informative patterns and updates the training set on the basis of a supervisor who attributes the labels to the selected unlabeled samples. This results in a representative training set that is as small as possible, thus reducing the cost of data labeling.

Many existing active learning techniques have focused on selecting the single most informative sample at each iteration [1], [2]. This can be inefficient, since the classifier has to be retrained for each new labeled sample. Thus, in this paper we focus on batch mode active learning, where a batch of $h > 1$ unlabeled samples is queried at each iteration by considering both uncertainty and diversity criteria [3], [4]. The uncertainty criterion is associated to the confidence of the supervised algorithm in correctly classifying the considered sample. Thus, the sample that has lowest confidence is the most uncertain. The diversity criterion aims at selecting a set of unlabeled samples that are as more diverse (distant one another) as possible in the feature space, thus reducing the redundancy among the samples selected at each iteration.

Active learning has been widely studied in the pattern recognition literature [1]–[3], [5]–[7]. Some pioneering works about the use of active learning for remote sensing image classification problems can be found in [8]–[12]. Mitra *et al.* [8] presented an active learning technique that selects n most uncertain samples, one from each OAA binary SVM. This is done choosing the sample closest to the current separating hyperplane of each binary SVM. In [9], an active learning technique is presented, which selects the unlabeled sample that maximizes the information gain between the a posteriori probability distribution estimated from the current training set and the training set obtained by including that sample into it. The information gain is measured by

the Kullback-Leibler divergence. In [10], Tuia *et al.* presented two batch mode active learning techniques for multiclass remote sensing image classification problems. The first extended the SVM margin sampling by incorporating diversity in kernel space, while the second is an entropy-based version of the query-by-bagging algorithm. In [11], Demir *et al.* investigated several SVM-based batch mode active learning techniques for the classification of remote sensing images. They also proposed a novel enhanced cluster-based diversity criterion in the kernel space. In [12], a fast cluster-assumption based active learning technique is presented, which considers only the uncertainty criterion and is particularly effective when the available initial training set is biased.

In this paper we propose a novel batch mode active learning technique for solving multiclass classification problems by using SVM classifiers with the OAA architecture. In the uncertainty step, the confidence of correct classification of each unlabeled sample is computed by defining a novel criterion function that considers both the smallest distance to the decision hyperplanes and the distances to the other hyperplanes of the binary SVMs for which the sample is within the margin. The regions that are within the margins of the binary SVMs are known as uncertain regions [10]. To select batch of most uncertain samples from all over these regions, the uncertain regions of the classifiers are split into multiple parts according to the number of decision boundaries passing on them. Then, the most uncertain samples are selected from each part. After selecting a batch of most uncertain/ambiguous samples, to minimize the redundancy and keep the diversity among these samples (this is important for reducing the size of the training set and thus the cost of labeling) we apply the kernel k-means clustering algorithm and extract from each cluster the representative sample that is closest to the corresponding cluster center (called medoid sample) for labeling.

To assess the effectiveness of the proposed method we compared it with three other batch mode active learning techniques existing in the literature using a hyperspectral and a multispectral remote sensing data sets.

The rest of this paper is organized as follows. The proposed SVM-based batch mode active learning technique is presented in Section II. Section III provides the description of the two remote sensing data sets used for experiments. Section IV presents different experimental results obtained on the considered data sets. Finally, Section V draws the conclusion of this work.

II. PROPOSED BATCH MODE ACTIVE LEARNING TECHNIQUE BASED ON MULTIPLE UNCERTAINTY

We propose a novel batch mode active learning technique for solving multiclass classification problems using the SVM classifier. Before presenting the proposed technique, we briefly recall the main concepts associated with it.

Let us assume that a training set consists of N labeled samples $(\mathbf{x}_i, y_i)_{i=1}^N$, where $\mathbf{x}_i \in \Re^d$ are the training samples and $y_i \in \{+1, -1\}$ are the associated labels (which represent classes ω_1 and ω_2). SVM is a binary classifier, which goal is to partition the d-dimensional feature space into two subspaces (one for each class) using a separating hyperplane. The training phase of the classifier can be formulated as an optimization problem by using the Lagrange optimization theory, which leads to the following dual representation:

$$\max_{\alpha} \quad \left\{ \sum_{i=1}^N \alpha_i - \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N y_i y_j \alpha_i \alpha_j K(\mathbf{x}_i, \mathbf{x}_j) \right\}$$

$$\sum_{i=1}^N y_i \alpha_i = 0$$

$$0 \leq \alpha_i \leq C$$

$$i = 1, 2, ..., N$$

where α_i are Lagrangian multipliers, $K(.,.)$ is a kernel function that implicitly models the classification problem into a higher dimensional space where linear separation between classes can be approximated and C is a regularization parameter that allows one to control the penalty assigned to training errors. The solution to the SVM learning problem is a global maximum of a convex function. The decision function $f(\mathbf{x})$ is defined as:

$$f(\mathbf{x}) = \sum_{\mathbf{x}_i \in SV} \alpha_i y_i K(\mathbf{x}_i, \mathbf{x}) + b \tag{1}$$

where SV represents the set of support vectors (the training pattern $\mathbf{x}_i$ is a support vector if the corresponding α_i has a nonzero value). For a given test sample $\mathbf{x}$, the sign of the discriminant function $f(\mathbf{x})$ defined in (1) is used to predict its class label.

In order to address multiclass problems on the basis of binary SVMs classifiers, the general approach consists of defining an ensemble of binary classifiers and combining them according

to some decision rules. The two most commonly adopted strategies are based on the OAA and one-against-one (OAO) strategies [13]. In this work, we adopt the OAA strategy, which involves a parallel architecture made up of n SVMs, one for each information class. Each SVMs solves a two-class problem defined by one information class against all the others. The reader is referred to [14], [15] for more details on the SVM approach.

In the next section we present a novel active learning technique that incorporates uncertainty and diversity criteria into two consecutive steps to select the h ($h > 1$) most informative unlabeled samples to be labeled at each iteration. The m ($m > h$) most uncertain samples which belong to different decision regions are selected in the uncertainty step, and the h less correlated samples among these m uncertain patterns are chosen in the diversity step.

A. Uncertainty Step

Many of the existing SVM-based active learning techniques for solving n ($n > 2$) class classification problem compute the confidence of correct classification of an unlabeled sample by considering the smallest distance to one of the n decision hyperplanes associated to the n binary SVM classifiers in an OAA architecture [4], [10], [11]. Then they select the m most uncertain samples from the unlabeled pool which have the lowest confidence. This kind of sampling method is known as marginal sampling (MS) and suffers two drawbacks when solving multiclass problems. First, it selects the most uncertain samples considering only the hyperplane closest to them without considering the position of the other $n - 1$ hyperplanes. Second, the m samples chosen by the uncertainty step might not be selected from all over the uncertainty regions. As a results, it needs more samples for converging. In this work we solve both problems by defining a novel criterion function and selecting the samples by taking into account the above-mentioned problems.

Let l be the number of samples in the unlabeled pool U, and $f_i(\mathbf{x}_j)(i = 1, \ldots, n; j = 1, \ldots, l)$ the decision value associated with the unlabeled pattern $\mathbf{x}_j$ from the ith binary SVM classifier. Initially each binary SVM classifier is trained with the few available initial labeled samples. After training, we get the n decision values $(f_1(\mathbf{x}_j), \ldots, f_n(\mathbf{x}_j))$ for the pattern $\mathbf{x}_j$ from the n binary SVM classifiers. Then for all $\mathbf{x}_j \in U$ we compute

$$Func(\mathbf{x}_j) = \{f_i(\mathbf{x}_j), f_i(\mathbf{x}_j) \in [-1, +1]\}. \tag{2}$$

$$C(\mathbf{x}_j) = cardinality\{Func(\mathbf{x}_j)\}. \tag{3}$$

$C(\mathbf{x}_j)$ counts the number of binary SVM classifiers in which $\mathbf{x}_j$ falls within the margin of the decision boundaries, i.e., $f_i(\mathbf{x}_j) \in [-1, +1]$. Thus, we can partition the uncertain regions of the classifiers into different parts depending on the value of $C(\mathbf{x})$ (see Fig.1). Note that this is an important feature of the proposed technique that helps us to select the most uncertain samples from diverse uncertain regions (see later). Then the uncertainty score $S(\mathbf{x}_j)$ of an unlabeled pattern $\mathbf{x}_j$ is computed as follows:

$$S(\mathbf{x}_j) = \min_{i=1,\ldots,n} \{|f_i(\mathbf{x}_j)|\} + \frac{\sum\limits_{f_i(\mathbf{x}_j) \in Func(\mathbf{x}_j)} |f_i(\mathbf{x}_j)|}{C(\mathbf{x}_j)}. \tag{4}$$
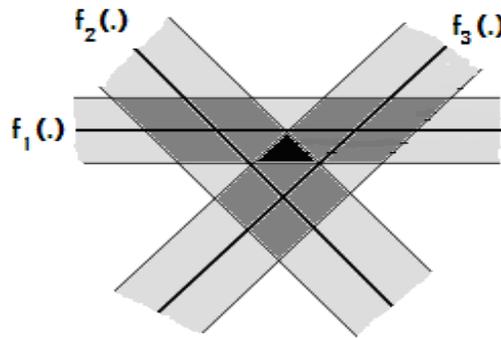
Fig. 1.   Illustrative examples of partition of the uncertainty regions of the classifiers into multiple parts depending on the number of geometrical margins of binary classifiers included in them. Different gray levels represent different uncertainty regions.

Equation (4) measures the confidence of correct classification of a pattern by considering both its distance to the closest hyperplane and its average distance to the hyperplanes for which it falls into the decision margins. In this way the uncertainty score does not depend only on the distance of the pattern to the most critical (closest) decision boundary but also on the average distance to all the other hyperplanes for which it is most uncertain. This allows us a more refined selection of most uncertain samples in the kernel space by considering multiple contributions to the uncertainty score, thus better capturing the properties of the multiclass problem in the evaluation of the uncertainty. If we select the m samples from the pool U which have minimum $S(.)$ value, it may happen that the samples are selected from a small portion of the uncertainty

regions. Thus, even if a diversity step is included in a later phase of the process, its effectiveness can be limited by the possible small diversity of the m selected samples (i.e., the diversity step does not become sufficiently effective to select diverse samples). As a result, the learning process may increase the number of samples necessary for converging. To avoid this problem, in our technique at each iteration of the active learning the m most uncertain samples are selected as shown in *algorithm 1*.

Algorithm 1 Algorithm for the selection of most uncertain samples

---

set $i = 0$

while $i \leq m$ **do**

   for $k = \max\limits_{j=1,\ldots,l}\{C(\mathbf{x}_j)\}$ to 1 step -1 **do**

     if $C(\mathbf{x}_j) = k$ and $S(\mathbf{x}_j) = \min\limits_{\substack{C(\mathbf{x}_t)=k \\ t=1,\ldots,l}}\{S(\mathbf{x}_t)\}$ **then**

       reset $C(\mathbf{x}_j) = 0$

       select $\mathbf{x}_j$ as an uncertain sample

       $i = i + 1$

     end if

   end for

end while

The uncertain regions of the classifier are divided into k sub-regions. These sub-regions are identified on the basis of the number of geometrical margins of binary SVMs included in them. Then, the algorithm first select k samples, one from each sub-region, according to the minimum uncertainty score. Then the process is repeated until the number m of most uncertain samples is obtained. Thus, the algorithm selects an almost equal number of most uncertain samples from each sub-region. This increases the probability to select the m samples from all over the uncertain regions of the classifiers.

B. Diversity Step

In this step a batch of h $(m > h > 1)$ samples that are diverse from each other are chosen among the m samples that are selected in the uncertainty step. To this end, we can apply any of the diversity criteria presented in the literature, e.g. angle based diversity [3], cluster based diversity [4], closest support vector based diversity [10], *etc.* to select a batch of diverse samples.

Here we prefer to use kernel k-means clustering as it works in the kernel space and already provided promising results in active learning [11]. We refer the reader to [11], [16] for details on the kernel k-means clustering technique applied to active learning. In the present work, first we apply the kernel k-means clustering algorithm to group the selected m most uncertain samples into h different clusters. Then, from each cluster the representative sample (medoid) is chosen for labeling. Thus, a total of h samples are selected, one from each cluster.

The process based on uncertainty and diversity steps is iterated until a stop criterion (which is related to the stability of the classification accuracy) is satisfied. *Algorithm 2* presents the complete procedure of the proposed technique.

III. DESCRIPTION OF DATA SETS

In order to assess the effectiveness of the proposed active learning technique, two data sets were used in the experiments. The first data set is made up of a hyperspectral image acquired on the forest of Paneveggio, near the city of Trento (northern Italy) on July, 2008. It consists of twelve partially overlapped images acquired by an AISA Eagle sensor in 126 bands ranging from 400m to 990m with spectral resolution of about 4.6nm and spatial resolution 1m. The size of the full image is 2199×2965 pixels. The available labeled samples were collected by ground survey. These samples were randomly split into a training set T of 4052 samples, and a test set TS (to compute the classification accuracy of the algorithms) of 2673 samples. In our experiments, first only few samples (2.5%) were randomly selected from T as initial training set L, and the rest were considered as unlabeled samples stored in the unlabeled pool U. Table I shows the land cover classes and the related number of samples used in the experiments.

The second data set is a Quickbird multispectral image acquired on the city of Pavia (northern Italy) on June, 2002. It has four pan-sharpened multispectral bands and a panchromatic channel with a spatial resolution of 0.7 m. The image size is 1024×1024 pixels. The available labeled samples were collected by photointerpretation. These samples were randomly split into a training set T of 5707 samples and a test set TS of 4502 samples. In our experiments, first only few samples (1.25%) were randomly selected from T as initial training set L, and the rest were stored in the unlabeled pool U. Table II shows the land-cover classes and the related number of samples used in the experiments.

Algorithm 2 Proposed SVM-based batch mode active learning technique based on multiple uncertainty

---

Train n binary SVMs by using the available labeled samples.

repeat

  for $j = 1$ to l step 1 **do**

    set $temp = NULL$

    set $k = 0$

    for $i = 1$ to n **do**

      if $f_i(\mathbf{x}_j) \in [-1, +1]$ **then**

        $k \leftarrow k + 1$

        $temp(k) = \mid f_i(\mathbf{x}_j) \mid$

      end if

    end for

    if $k > 0$ **then**

      $C(\mathbf{x}_j) = k$

$$S(\mathbf{x}_j) = \min_{i=1,\ldots,k}\{temp(i)\} + \frac{\sum\limits_{i=1}^{k} temp(i)}{C(\mathbf{x}_j)}$$

    else

      $C(\mathbf{x}_j) = 0$

    end if

  end for

  Call *algorithm 1* to select m most uncertain samples

  Apply kernel k-means clustering algorithm to the selected m samples fixing $k = h$.

  Select the representative samples from each of the h clusters.

  Assign true labels to the h selected samples and update the training set.

  Retrain the n binary SVM classifiers by using the updated training set.

until the stop criterion is satisfied.

IV. EXPERIMENTAL RESULTS

A. Design of experiments

In our experiments we adopted an SVM classifier with RBF kernel. The SVM parameters $\{\sigma, C\}$ were derived by applying the cross-validation technique. C is a parameter controlling the tradeoff between model complexity and training error, while σ is the spread of the Gaussian

TABLE I

NUMBER OF SAMPLES FOR EACH CLASS IN THE INITIAL TRAINING SET(L), IN THE TEST SET(TS) AND
IN THE UNLABELED POOL(U) FOR THE HYPERSPECTRAL DATA SET

Classes	**L**	**TS**	**U**
Picea Abies	39	1135	1515
Larix Decidua	13	308	520
Pinus Mugo	6	160	234
Alnus Viridis	3	70	122
No Forest	40	1000	1560
Total	101	2673	3951

TABLE II

NUMBER OF SAMPLES FOR EACH CLASS IN THE INITIAL TRAINING SET(L), IN THE TEST SET(TS) AND
IN THE UNLABELED POOL(U) FOR THE MULTISPECTRAL DATA SET

Classes	**L**	**TS**	**U**
Water	2	215	178
Tree areas	4	391	344
Grass areas	4	321	319
Road	12	613	975
Shadow	9	666	709
Red building	29	1620	2267
Gray building	7	427	590
White building	3	249	255
Total	70	4502	5637

kernel. The cross-validation procedure aims at selecting the best values for the parameters of the initial SVM. The same RBF kernel function is also used to implement the kernel k-means algorithm.

To assess the effectiveness of the proposed technique we compared it with three other methods: i) the random sampling (RS); ii) the marginal sampling with closest support vector (MS-cSV) [10]; and iii) the marginal sampling with kernel k-means clustering (MS-KKC). In the RS approach, at each iteration a batch of h samples are randomly selected from the unlabeled pool U and included into the training set. The MS-cSV approach used MS to select the m most uncertain samples i.e., it selects the m $(m > h)$ samples that have the smallest distance to one

of the n decision hyperplanes associated to the n binary SVMs in an OAA architecture. Then, the h most informative samples that do not share the closest support vector are added to the training set. In order to assess the effectiveness of all the components of the proposed technique, we also carried out experiments by using the MS criterion in the uncertainty step to select the m most uncertain samples and then exploiting exactly the same diversity criterion as used by the proposed method (i.e., the kernel k-means clustering) to select the h most informative samples in the diversity step. In this way we can appreciate the gain provided by both the proposed uncertainty and diversity criteria. We call this algorithm MS-KKC. Note that in the present experiments, the value of m is fixed to $m = 3h$ for a fair comparison among the different techniques.

The multiclass SVM with the standard OAA architecture has been manually implemented by using the LIBSVM library (for Matlab interface) [17]. All the active learning algorithms presented in this paper and also the kernel k-means algorithm have been implemented in Matlab.

B. Results

In order to understand the effectiveness of the proposed technique, in the first experiment we compared the performance of the proposed method with those of the three techniques described in the previous subsection. For the hyperspectral data set, initially only 101 labeled samples were included in the training set and 20 samples were selected at each iteration of active learning. The whole process was iterated 44 times resulting in 981 samples in the training set at convergence. For the multispectral data set, initially only 70 labeled samples were included in the training set and 20 samples were selected at each iteration of active learning. The whole process was iterated 19 times resulting in 450 samples in the training set at convergence. The active learning process was repeated for 20 trials to reduce the random effect on the results. Figs. 2 and 3 show the average overall classification accuracies and standard deviations provided by different methods versus the number of samples included in the training set at different iterations for the hyperspectral and the multispectral data sets, respectively. From these figures, one can see that the proposed active learning technique always resulted in higher classification accuracy than the other techniques. Furthermore, for the hyperspectral data set, the proposed technique yielded an accuracy of 90.87% with only 981 labeled samples, while using the full pool as training set (i.e., 4052 samples) we obtained an accuracy of 91.37%. For the multispectral data set, the

proposed technique yielded an accuracy of 86.62% with only 450 labeled samples, while using the full pool as training set (i.e., 5707 samples) we obtained an accuracy of 86.80%. It is worth noting that both the proposed uncertainty and diversity criteria are effective, as proven by the improvement obtained by the proposed method with respect to the use of only the kernel k-means clustering and the standard MS (MS-KCC). The analysis of the behavior of the standard deviation of the accuracy on the different trials versus the number of considered samples (see Figs. 2(b) and 3(b)) also points out that on both data sets the proposed technique has a smaller standard deviation than (or in some cases comparable with) the other algorithms. This confirms the good stability of the proposed method versus the choice of initial training samples.

(a)

(b)
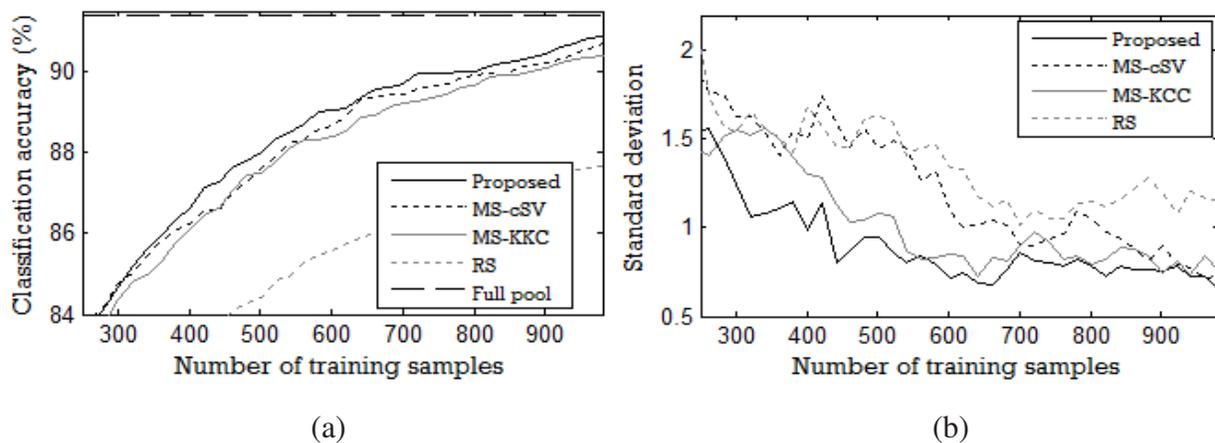
Fig. 2. (a) classification accuracy and (b) standard deviation of the accuracy over twenty runs provided by the Proposed, the MS-cSV, the MS-KKC, and the RS methods for the hyperspectral data set.

The second experiment was devoted to analyze the usefulness of both terms used in (4). To this end, we computed the uncertainty score of each unlabeled pattern in three different ways: i) considering only the first term of (4) (i.e., the minimum distance term), ii) considering only the second term of (4) (i.e., the average distance term), and iii) considering both terms as used in the proposed method. Figs. 4(a) and 4(b) show the average classification accuracy gain obtained by considering both terms of (4) over considering only the minimum and only the average distance term separately, for the hyperspectral and multispectral data sets, respectively. From the analysis of these figures one can conclude that the uncertainty score computed by using both terms is able to choose more meaningful samples than by using them separately. It is worth noting that even if the increase of accuracy is limited on the considered data sets, we have improvements at all

(a)   (b)
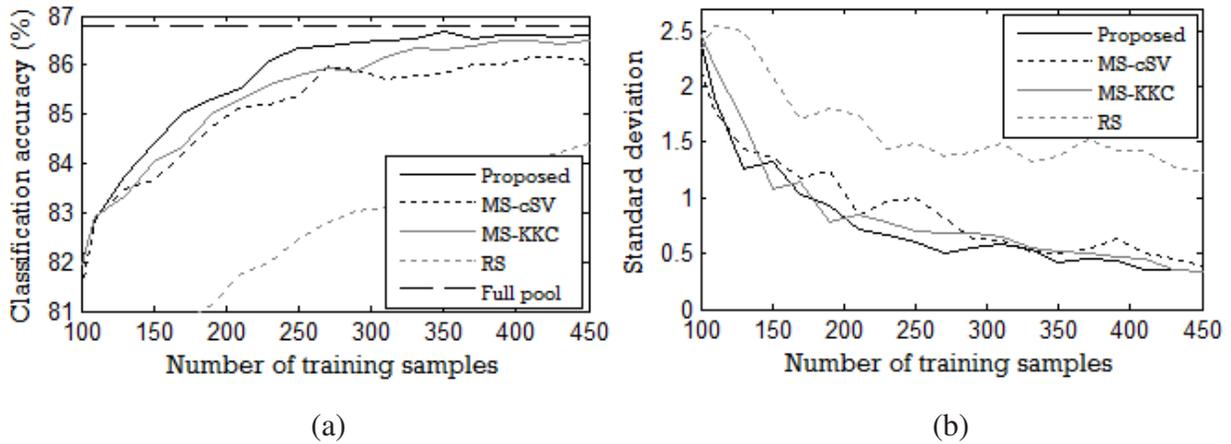
Fig. 3.   (a) classification accuracy and (b) standard deviation of the accuracy over twenty runs provided by the Proposed, the MS-cSV, the MS-KKC, and the RS methods for the multispectral data set.

iterations of the active learning process and the use of the two terms does not affect significantly the computation time taken from the method.

(a)   (b)
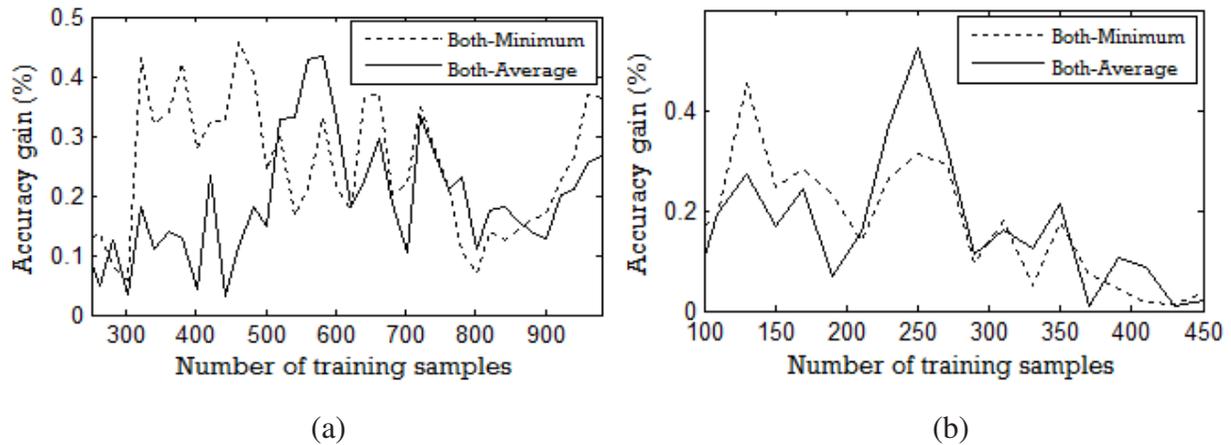
Fig. 4.   Average classification accuracy gain obtained by considering both terms of (4) over considering only the minimum and only the average distance terms for the (a) hyperspectral and (b) multispectral data sets.

The last experiment shows the effectiveness of the different techniques in terms of computational load. All the experiments were carried out on a PC (INTEL(R) Core(TM)2 Duo 2.0 GHz with 2.0 GB of RAM) with the experimental setting (i.e., number of initial training samples, batch size, iteration number, etc.) described in the previous experiment. Fig. 5 shows the computational time (in seconds) versus the number of training samples (i.e., of iterations) required by the

proposed, the MS-cSV, the MS-KKC, and the RS techniques for the multispectral data set. From this figure, one can see that the computational time required by the proposed method is almost similar to the computational time taken by the MS-KKC approach and the computational time taken by the MS-cSV techniques is much higher compared to the proposed method. Indeed, when the number of training samples increases (i.e., the number of SVs increases) the MS-cSV technique takes much time to find out the uncertain samples which are closest to the distinct SVs. The RS method was obviously the most efficient in terms of computational load. Nonetheless, it resulted in the lowest classification accuracy. Similar observations were found on the hyperspectral data set (results are not reported for space constraints).
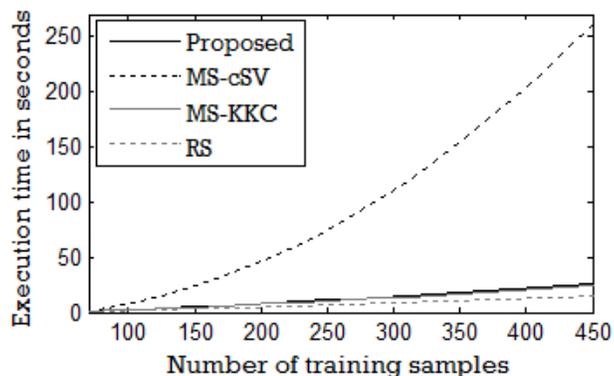
Fig. 5.   Computational time taken by the Proposed, the MS-cSV, the MS-KKC, and the RS techniques at each iteration for the multispectral data set.

On the basis of the all the above-mentioned experiments, we can conclude that the proposed technique: i) required a smaller number of labeled samples for converging than the other methods; ii) is robust to the choice of initial training samples; and iii) is efficient in terms of computational complexity.

V. CONCLUSION

In this paper we have presented a novel batch mode active learning technique for solving multiclass problems with SVM classifiers, which considers both uncertainty and diversity criteria. The uncertainty of each unlabeled sample is measured by defining a criterion that, unlike literature methods, considers the distance of the sample from all the hyperplanes of the binary SVMs in the multiclass OAA architecture for which the sample is within the margin (multiple uncertainty).

Then to select the m most uncertain samples which are close to different decision hyperplanes and belong to all over the decision region, the uncertainty regions of the classifiers are divided into multiple parts. The partition is done by taking into account the number of binary SVMs for which the considered area is uncertain (i.e., the number of SVMs that have the margin falling into the considered part of the kernel space). To minimize the redundancy and keep the diversity among these samples, the kernel k-means clustering algorithm is applied. Then a representative sample (medoid) from each cluster is selected for labeling.

To empirically assess the effectiveness of the proposed method we compared it with other three batch mode active learning techniques using hyperspectral and multispectral remote sensing data sets. In this comparison we observed that the proposed method provided higher accuracies with improved stability than those achieved by some of the most effective techniques presented in the literature. More in general, compared with existing methods, the proposed technique provided the best trade-off among classification accuracy, number of labeled samples needed to reach the convergence, and computational complexity. As future developments of this work, we plan to both extend the experimental comparison to other methods and to use the concepts introduced in this paper with the OAO architecture for multiclass SVM classifiers.

ACKNOWLEDGMENTS

REFERENCES

[1] D. A. Cohn, Z. Ghahramani, and M. I. Jordan, "Active learning with statistical models," *J. Artificial Intelligence Research*, vol. 4, no. 1, pp. 129–145, 1996.

[2] S. Tong and D. Koller, "Support vector machine active learning with applications to text classification," *J. Machine Learning Research*, vol. 2, no. 1, pp. 45–66, 2002.

[3] K. Brinker, "Incorporating diversity in active learning with support vector machines," in *Proc. 20th ICML*, 2003, pp. 59–66.

[4] R. Liu, Y. Wang, T. Baba, D. Masumoto, and S. Nagata, "SVM-based active feedback in image retrieval using clustering and unlabeled data," *Pattern Recognition*, vol. 41, pp. 2645–2655, 2008.

[5] C. Campbell, N. Cristianini, and A. Smola, "Query learning with large margin classifiers," in *Proc. 17th ICML*, 2000, pp. 111–118.

[6] P. Mitra, C. A. Murthy, and S. K. Pal, "A probabilistic active support vector learning algorithm," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 26, no. 3, pp. 413–418, 2004.

[7] M. Li and I. K. Sethi, "Confidence-based active learning," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 28, no. 8, pp. 1251–1261, 2006.

[8] P. Mitra, B. U. Shankar, and S. K. Pal, "Segmentation of multispectral remote sensing images using active support vector machines," *Pattern Recognition Letters*, vol. 25, no. 9, pp. 1067–1074, 2004.

[9] S. Rajan, J. Ghosh, and M. M. Crawford, "An active learning approach to hyperspectral data classification," *IEEE Trans. Geoscience and Remote Sensing*, vol. 46, no. 4, pp. 1231–1242, 2008.

[10] D. Tuia, F. Ratle, F. Pacifici, M. F. Kanevski, and W. J. Emery, "Active learning methods for remote sensing image classification," *IEEE Trans. Geoscience and Remote Sensing*, vol. 47, no. 7, pp. 2218–2232, 2009.

[11] B. Demir, C. Persello, and L. Bruzzone, "Batch-mode active-learning methods for the interactive classification of remote sensing images," *IEEE Trans. Geoscience and Remote Sensing*, vol. 49, no. 3, pp. 1014–1031, 2011.

[12] S. Patra and L. Bruzzone, "A fast cluster-assumption based active learning technique for classification of remote sensing images," *IEEE Trans. Geoscience and Remote Sensing*, vol. 49, no. 5, pp. 1617–1626, 2011.

[13] F. Melgani and L. Bruzzone, "Classification of hyperspectral remote sensing images with support vector machines," *IEEE Trans. Geoscience and Remote Sensing*, vol. 42, no. 8, pp. 1778–1790, 2004.

[14] B. E. Boser, I. M. Guyon, and V. N. Vapnik, "A training algorithm for optimal margin classifiers," in *Proc. 5th Annual Workshop Computational Learning Theory*, 1992, pp. 144–152.

[15] C. J. C. Burges, "A tutorial on support vector machines for pattern recognition," *Data Mining and Knowledge Discovery*, vol. 2, no. 2, pp. 121–167, 1998.

[16] R. Zhang and A. I. Rudnicky, "A large scale clustering scheme for kernel k-means," in *Proc. 16th ICPR*, 2002, pp. 289–292.

[17] C.-C. Chang and C.-J. Lin, *LIBSVM: a library for support vector machine*, 2001, software available at http://csie.ntu.tw/~cjlin/libsvm.