Title: A Cluster-Assumption Based Batch Mode Active Learning Technique

This paper appear in: Pattern Recognition Letters

Date of Publication: 2012

Author(s): Swarnajyoti Patra and Lorenzo Bruzzone

Volume: 33

Page(s): 1042-1048

DOI: 10.1016/j.patrec.2012.01.015

A Cluster-Assumption Based Batch Mode Active Learning Technique

Swarnajyoti Patra and Lorenzo Bruzzone *

Department of Information Engineering and Computer Science, University of Trento, Via Sommarive 14, I-38123 Trento, Italy

Abstract

In this paper, we propose an active learning technique for solving multiclass problems with support vector machine (SVM) classifiers. The technique is based on both uncertainty and diversity criteria. The uncertainty criterion is implemented by analyzing the one-dimensional output space of the SVM classifier. A simple histogram thresholding algorithm is used to find out the low density region in the SVM output space to identify the most uncertain samples. Then the diversity criterion exploits the kernel k-means clustering algorithm to select uncorrelated informative samples among the selected uncertain samples. To assess the effectiveness of the proposed method we compared it with other batch mode active learning techniques presented in the literature using one toy data set and three real data sets. Experimental results confirmed that the proposed technique provided a very good tradeoff among robustness to biased initial training samples, classification accuracy, computational complexity, and number of new labeled samples necessary to reach the convergence.

Key words: Active learning, cluster assumption, entropy, query function, support vector machine

1 Introduction

In the machine learning literature many supervised algorithms have been proposed to perform the pattern classification tasks. In all these methods, the classification accuracy relies on the quality of the labeled samples used in the learning phase. However, in many problems the collection of labeled samples that can precisely represent the statistics of all the classes is a time consuming and complex process. Redundant samples are often included in the training

* Corresponding author. Email:lorenzo.bruzzone@ing.unitn.it

set, thus slowing down the training process of the classifier without improving classification results. In order to reduce the time and cost of labeling, the samples in the training set should not be redundant and should contain the maximum amount of information for the discrimination of classes. Active learning is an effective approach to the solution of this kind of problem. The learning process repeatedly queries unlabeled samples to select the most informative patterns to be labeled and updates the training set on the basis of a supervisor who attributes the labels to the selected unlabeled samples.

Many existing active learning methods select informative samples by considering only an uncertainty criterion [1–4]. Depending on the criterion considered, at each iteration either i) single [1,2] or ii) multiple [3,4] uncertain samples can be labeled. The first approach can be inefficient since the classifier has to be retrained for each new labeled sample added to the training set. The second approach can be inefficient too since there might exist redundancy between the selected samples. Some active learning techniques exist that query a batch of unlabeled samples at each iteration by considering both uncertainty and diversity criteria [5–7]. The uncertainty criterion is associated to the confidence of the supervised algorithm in correctly classifying the considered sample, while the diversity criterion aims at selecting a set of unlabeled samples that are as more diverse as possible in the feature space, thus reducing the redundancy among the samples selected at each iteration. The combination of the two criteria results in the selection of the potentially most informative set of samples at each iteration of the active learning process.

In this article we propose a novel active learning technique for solving multiclass classification problems with SVM classifiers by considering both uncertainty and diversity criteria. The main idea of the technique is to exploit the cluster assumption which states that the decision boundary between classes has to lie in the low-density regions of the feature space [8]. Initially each binary SVM classifier is trained with a small number of labeled samples. After training, an histogram corresponding to each binary SVM is constructed in the one-dimensional output space of the classifier by considering the output scores of the unlabeled samples in [-1, +1]. Since the classifier ranks each sample from the most likely members to the most unlikely members of a class, the samples whose output scores fall in the valley region of the histograms are the most uncertain. Thus, we can identify the uncertain samples by finding a threshold corresponding to the valley region in each histogram. Then a batch of samples is selected from the unlabeled pool whose output scores are closest to one of the selected threshold values. After selecting a batch of uncertain samples, to minimize the redundancy and keep the diversity among these samples, we apply kernel k-means clustering algorithm and query the sample from each cluster whose output score is most uncertain. Since the proposed technique selects the unlabeled samples from low-density regions in the kernel space, it is not strongly affected by the set of initial training samples and by the previous training results, thus allowing relatively fast convergence also by starting with biased (poor) initial training samples.

The rest of this paper is organized as follows. Section 2 describes the active learning process and briefly surveys existing active learning methods. The proposed cluster assumption based active learning approach is presented in Section 3. Section 4 provides the description of the four data sets used for experiments. Section 5 presents different experimental results obtained on the considered data sets. Finally, Section 6 draws the conclusion of this work.

2 Active learning

A general active learner can be modeled as a quintuple (G, Q, S, L, U) [4]. Initially, the training set L has few labeled samples to train the classifier G. After that, the query function Q is used to select a set of most informative samples from the unlabeled pool U and the supervisor S assigns a class label to each of them. Then, these new labeled samples are included into L and the classifier G is retrained using the updated training set. The closed loop of querying and retraining continues for some predefined iterations or until a stop criterion is satisfied.

The query function is fundamental in the active learning process. Several methods have been proposed in the literature that differs only in their query functions. In [9], Fukumizu has proposed a statistical active learning approach to train multilayer perceptron for performing regression. Roy and McCallum [10] proposed an active learning method that select the next unlabeled sample to be labeled on the basis of the minimization of the future error rate which is estimated by using two different techniques. One of the most popular active learning technique based on SVM is to select the data point closest to the current separating hyperplane, which is also referred to as marginal sampling (MS) [1]. An active learning strategy based on version space splitting is presented in [2]. Another class of active learning methods is based on query-by-committee [11,12], wherein the sample that has highest disagreement among the committee of classifiers is chosen for the labeling.

It is worth noting that all of the above-mentioned methods consider only one sample at each iteration. However, in many problems it is necessary to speed up the learning process by selecting more than one sample at each iteration. In [3], Mitra *et al.* have presented a probabilistic active learning approach, wherein query samples are selected according to both the distance from the current separating hyperplane and a confidence factor estimated from a set of test samples using the nearest neighbor technique. In [4], an approach is proposed that estimates the uncertainty level of each sample according to the output score of a classifier and selects only those samples whose output scores are within the uncertainty range. In [13], we proposed a fast cluster assumption based active learning technique that also work on the one-dimensional output space of the SVM classifier to solve multiclass classification problems. Note that all the aforementioned methods selects batch of sample at each iteration by considering only the uncertainty criterion. This may result in the selection of redundant samples which reduce the speed of the classifier without adding any additional information. In order to address this shortcoming, Brinker [5] has presented a method that selects batch of samples by incorporating a diversity measure that considers the angles between the induced classification hyperplane. Clustering based diversity measures that are incorporated to design the query functions of the active learning are discussed in [6,14]. In [7], two batch mode active learning techniques for multiclass remote sensing image classification problems are proposed (more details on these approaches that will be used as benchmark in our experimental analysis are given in the experiment section).

3 Proposed method

Here we present a cluster assumption based batch mode active learning technique for solving multiclass classification problems with SVM classifiers. Before presenting the proposed technique, we briefly recall the main concepts associated with SVM classifiers. The reader is referred to [15] for more details on the SVM approach.

Let us assume that a training set consists of N labeled samples $(x_i, y_i)_{i=1}^{N}$, where $x_i \in \Re^d$ are the training samples and $y_i \in \{+1, -1\}$ are the associated labels (which model classes ω_1 and ω_2). SVM is a binary classifier, whose goal is to divide the d-dimensional feature space into two subspaces (one for each class) using a separating hyperplane. An interesting feature of SVM is related to the possibility to implicitly project the original data into a higher dimensional feature space via a kernel function $K(.,.)$, which satisfies the Mercers conditions [15]. The solution to the SVM learning problem is a global maximum of a convex function. The decision function $f(x)$ is defined as:

$$f(x) = \sum_{x_i \in SV} \alpha_i y_i K(x_i, x) + b \qquad (1)$$

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

In order to address multiclass problems on the basis of binary SVMs classifiers, in this work, we adopt the one-against-all (OAA) strategy, which involves a parallel architecture made up of n SVMs, one for each information class. Each SVM solves a two-class problem defined by one information class against all the others [16].

In the following section we propose an active learning technique that incorporate uncertainty and diversity criteria in two consecutive steps to select the h ($h > 1$) most informative unlabeled samples to be labeled at each iteration. The novelty of the proposed technique consists in the adopted uncertainty and diversity criteria.

3.1 Uncertainty step

The proposed uncertainty criterion is developed under the hypothesis that for the considered data sets the cluster assumption holds. Initially each binary SVM classifier is trained with the few available labeled samples. After training, an histogram corresponding to each classifier is constructed in the one-dimensional output space of the classifier by considering the output scores of the samples in [-1, +1]. In the histogram, the region of interest is quantized into N mutually exclusive intervals called bins. We assume that all bins have an equal width (uniform quantization). The probability to have the output in a given bin is given by the number of samples whose output scores fall in that bin divided by the total number of samples in the histogram (i.e., the samples given as input to the classifier). Since the classifier ranks samples from the most likely members to the most unlikely members of a class, according to the cluster assumption (the decision boundary has to lie in the low density regions) the samples whose output score fall in the valley region of the histogram are the most uncertain. Thus we can work in the one-dimensional output space of the classifier to identify the uncertain samples by defining a threshold on the histogram which is passing through this valley region. This simple strategy avoids the complexity of the design of the query function in the original features space, which may be associated with complicated decision regions and thus can become computationally demanding. To detect a proper threshold on the histogram any thresholding technique existing in the pattern recognition literature can be used [17]. For the present work, without loosing in generality, we applied the Kapur's entropy based histogram thresholding technique [13,18].

In greater detail, let us consider a problem with n classes. n binary SVMs are initially trained with the current training set and the functional distance $f_i(x)$ of each i^{th} binary SVM ($i = 1, ..., n$) is calculated for all the unlabeled samples $x \in U$ and the related histogram H_i is generated by considering the output score value in [-1, +1]. Thus each binary SVMs classifier generates a separate histogram considering its output score values. Then a threshold t_i from the histogram H_i is detected by applying the histogram based thresholding method. After finding n thresholds, the uncertainty of a sample $x \in U$ is computed as follows:

$$c(x) = \min_{i=1,\dots,n} \{|f_i(x) - t_i|\} \tag{2}$$

Now the m most uncertain samples from U are selected which have minimum c(.) value. Note that, unlike in [13] where a fixed number of uncertain samples are selected by considering each individual binary SVM, here the uncertain samples are selected by considering all the binary SVM together. This allows us to select variable numbers of samples from uncertain regions associated with different SVMs.

3.2 Diversity step

In this step a batch of h $(m > h > 1)$ samples which are diverse from each other are chosen among the m samples that are selected in the uncertainty step. In the present work, first we apply the kernel k-means clustering algorithm [19] to divide the selected m uncertain samples into h different clusters. Then, from each cluster the most uncertain sample is chosen. The amount of uncertainty is evaluated according to the criterion proposed in the uncertainty step. This is different from the strategy used in the standard cluster based technique which usually selects the barycenter of the cluster as the representative sample of that cluster.

In greater detail, let us assume that the kernel k-means clustering algorithm divides the m samples into $k = h$ clusters $C_1, C_2, ..., C_h$ in the kernel space. After $C_1, C_2, ..., C_h$ are obtained, the h most informative samples are selected as

$$x_k = \arg \min_{x \in C_k} \left\{ \min_{i=1,\dots,n} \{|f_i(x) - t_i|\} \right\}, \quad k = 1, 2, ..., h \tag{3}$$

where x_k is the kth sample selected using the proposed query function and corresponds to the most uncertain sample of the cluster C_k. Thus, a total of h samples (one for each cluster) are selected . The process is iterated until a stop criterion (which is related to the stability of the classification accuracy) is satisfied. *Algorithm 1* presents the complete procedure of the proposed technique.

4 Data sets description

In order to assess the effectiveness of the proposed active learning technique, four data sets were used in the experiments. The first one is a toy data set which is made up of three linearly separable classes as shown in Fig. 1(a). It contains 43 samples, and only 3 samples (one from each class) are chosen as initial training samples; the remaining 40 samples are in the unlabeled pool U. This data set is used for illustrating the properties of the proposed technique.

The second UCI ionosphere data set is a well-known and widely used benchmark for pattern classification tasks [20]. These radar data were collected by a system in Goose Bay, Labrador. It consists of 351 patterns (free electrons in the ionosphere). There are 34 continuous input features and two overlapping classes.

The third is a more complicated vowel data set widely used for pattern classification tasks [21]. It consists of 871 patterns (Indian Telugu vowel sounds). It has three integers input features with six overlapping classes.

Algorithm 1: Proposed cluster-assumption based batch mode active learning technique

Step 1: Train n binary SVMs by using an initial small number of labeled samples. Let $f_i(.)$ be the decision function of the i^{th} binary SVM classifier.
Repeat
  For i=1 to n
    Step 2: For the i^{th} binary SVM classifier generates the corresponding histogram H_i by considering the output scores of the unlabeled samples $x \in U$ whose output value $f_i(x) \in [-1, +1]$.
    Step 3: Detect the threshold t_i from the histogram H_i by using the entropy based histogram thresholding technique.
  End for
  Step 4: Select m ($m > h$) samples which have minimum $c(.)$ value defined in (2).
  Step 5: Apply kernel k-means clustering algorithm to the m samples selected in step 4 fixing $k = h$.
  Step 6: Select one sample from each of the h clusters using (3).
  Step 9: Assign true labels to the h selected samples and update the training set.
  Step 10: Retrain the n binary SVMs by using the updated training set.
Until the stop criterion is satisfied.

The last data set is a Quickbird multispectral remote sensing image acquired on the city of Pavia (northern Italy) on June, 2002. It consists four pansharpened multispectral bands and a panchromatic channel with a spatial resolution of 0.7m. The size of the full image is 1024×1024 pixels and there are eight classes. The available labeled samples were collected by photointerpretation.

For all data sets, the available samples were randomly split into a training set T and a test set TS. First, only few samples were randomly selected from T as initial training set L, and the rest were stored in the unlabeled pool U. Table 1 shows all the classes and the related number of samples used in the experiments for the ionosphere, the vowel and the remote sensing data sets.

Table 1
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 ionosphere, the vowel and the remote sensing data sets

Data sets	**Classes**	**Number of samples**		
		L	**TS**	**U**
ionosphere	good	11	67	147
	bad	6	38	82
	Total	17	105	229
vowel	∂	4	22	46
	a	4	27	58
	e	8	52	112
	i	7	45	99
	o	10	62	135
	u	9	54	117
	Total	42	262	567
remote sensing	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

5 Experimental results

5.1 Design of experiments

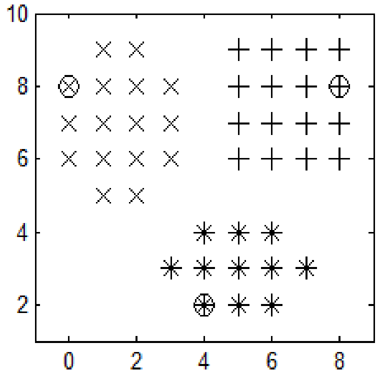
In our experiments we adopted an SVM classifier with RBF kernel. The SVM parameters $\{\sigma, C\}$ were derived by applying the cross-validation technique. The RBF kernel is also used to implement the kernel k-means algorithm.

To assess the effectiveness of the proposed technique we compared it with four other methods: i) the Brinker's method [5], which is based on marginal sampling with angle based diversity (MS-ABD); ii) the marginal sampling with closest support vector diversity (MS-cSV) [7]; iii) the entropy query-by-bagging (EQB) [7]; and iv) the random sampling (RS). For multiclass problems, the MS-ABD approach first uses 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 diverse samples among the m samples are chosen by applying the angle based diversity as presented in [5]. Here we set the value of the weighting parameter that tunes the tradeoff between uncertainty and diversity at $\lambda = 0.5$. The MS-cSV considers the MS criterion to select the m most uncertain samples. Then, the h most uncertain samples which do not share the same closest support vector are added to the training set. The EQB selects the h most uncertain samples according to the maximum disagreement between a committee of classifiers. The results of the EQB are obtained by fixing the number of predictors to eight and selecting bootstrap samples containing 75% of initial training patterns. 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. Note that, in the present experiments, the value of m is fixed to $m = 3h$ for a fair comparison among the different techniques.

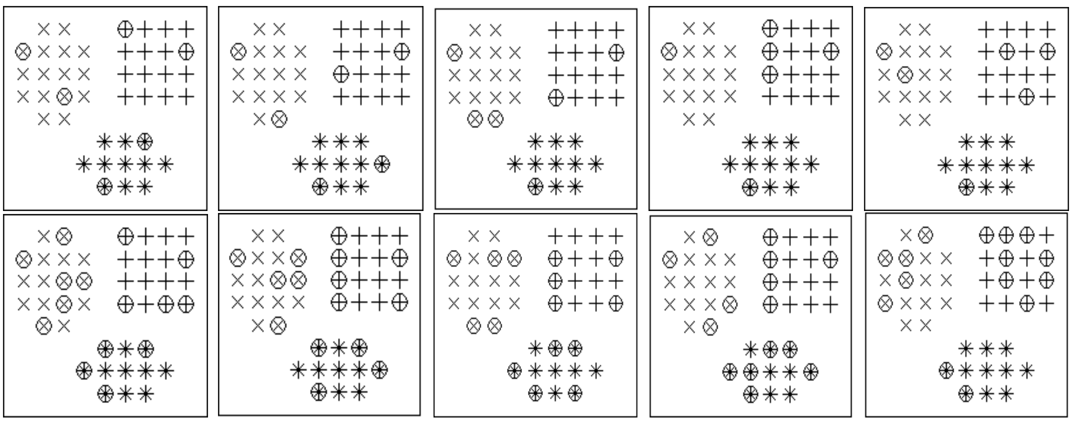
The multiclass SVM with the standard OAA architecture has been implemented by using the LIBSVM library (for Matlab interface) [22]. All the active learning algorithms presented in this paper have been implemented in Matlab.

5.2 Results

In order to understand the potential of the proposed technique, in the first experiment we compared the different active learning methods by using the toy data set described in the previous subsection. Initially, only three samples (one from each class) are chosen for the training (see Fig. 1(a)) and 3 additional samples are selected at each iteration of active learning. The process is iterated 4 times to have 15 samples in the training set at the end. Fig. 1 shows the unlabeled samples (represented with circles) which are selected by different active learning methods after the end of the 1st and the 4th iteration. From this figure one can see that, for example, according to the use of the cluster assumption at the initial stage of the training, the proposed technique selects samples more representative of the general problem than the other techniques. For a quantitative analysis, Table 2 reports the classification accuracy obtained by the proposed, the MS-ABD, the MS-cSV, the EQB and the RS methods at different iterations. From the table one can see that the proposed technique obtained 100% classification accuracy after the 1st iteration (i.e., by using only 6 labeled samples), while the other most effective techniques (i.e., the MS-ABD, the MS-cSV and the EQB) needed at least 2 iterations (i.e., 9 samples) to achieve the same accuracy. In other words, although this is a simple example, starting from a suboptimal data set, the proposed technique, thanks to the low-density criterion, reaches the convergence decreasing of 33% the number of new labeled samples with respect to the other literature methods.



(a)



(b) (c) (d) (e) (f)

Fig. 1. (a) Toy data set with initial labeled samples represented with circles. The samples represented with circles selected by (b) the Proposed; (c) the MS-ABD; (d) the MS-cSV; (e) the EQB; and (f) the RS methods after the 1st (upper part of the figure) and the 4th (lower part of the figure) iterations.

The second experiment was carried out to compare the performance of the

Table 2
Overall classification accuracy ($\overline{OA}$) produced by the different techniques at different iterations (toy data set)

Itr	Training	$\overline{OA}$				
No	Samples	Proposed	MS-ABD	MS-cSV	EQB	RS
0	3	97.43	97.43	97.43	97.43	97.43
1	6	100	97.43	94.87	89.74	92.30
2	9	100	100	100	100	94.87
3	12	100	100	100	100	92.30
4	15	100	100	100	100	97.43

proposed method with those of the four techniques described in the previous subsection on the three real data sets considered in this paper. For the ionosphere and the vowel data sets, initially only 17 and 42 labeled samples were included in the training set, respectively, and 5 samples were selected at each iteration of active learning. The whole active learning process was iterated 10 times resulting in 67 and 92 samples in the training sets at convergence. For the remote sensing 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. For all the three data sets, the active learning process was repeated for 20 trials with 20 different initial training sets (generated randomly) to reduce the random effect on the results. Figs. 2(a), (b), and (c) show the average overall classification accuracies provided by different methods versus the number of samples included in the training set at different iterations for the ionosphere, the vowel and the remote sensing data sets, respectively. From these figures, one can see that the proposed active learning technique always resulted in better (or comparable) classification accuracy than the other techniques. For a quantitative analysis, Table 3 report the mean($\overline{OA}$), and standard deviation (s) of the overall accuracy, as well as the average kappa (k) accuracies obtained on 20 runs at three different iterations. From the table, one can see that the standard deviation of the proposed approach is always smaller than those of the other techniques. This confirms the better stability of the proposed method versus the choice of initial training samples.

Most of the active learning approaches select the uncertain samples depending on the current decision hyperplane. If the initial training samples are biased, i.e., they do not provide precise representation of the classification problem, then they may fail to select proper informative samples at the initial stage of learning. This results in a slowing down of the convergence process. On the contrary, the proposed technique selects the uncertain samples from the low-density region in the classifier output space and thus it is less dependent on

Table 3
Average overall classification accuracy ($\overline{OA}$), its standard deviation (s) and average kappa (k) accuracy obtained on twenty runs for different training data size of the ionosphere, the vowel, and the remote sensing data sets

Data sets	$\|L\|$	**Proposed**			**MS-ABD**			**MS-cSV**			**EQB**			**RS**		
		$\overline{OA}$	s	k	$\overline{OA}$	s	k	$\overline{OA}$	s	k	$\overline{OA}$	s	k	$\overline{OA}$	s	k
ionosphere	37	**92.20**	**1.48**	**.826**	91.41	2.42	.804	90.78	2.67	.790	89.90	2.36	.772	88.55	3.07	.739
	52	**94.10**	**1.30**	**.868**	94.08	1.47	**.868**	93.97	1.92	.865	93.26	1.36	.850	90.20	2.80	.777
	67	**95.28**	**0.87**	**.895**	95.25	.91	.894	95.24	1.21	.894	94.93	0.97	.887	91.47	2.39	.808
vowel	62	**79.24**	**2.65**	**.745**	78.47	3.45	.736	78.44	3.77	.735	77.92	2.98	.729	76.92	3.48	.717
	77	80.95	**2.53**	.766	80.90	2.83	.766	**81.03**	3.11	**.767**	80.03	3.06	.754	78.36	3.16	.735
	92	**82.81**	**2.37**	**.789**	81.99	2.51	.779	82.61	2.47	.786	81.80	2.73	.776	79.77	2.37	.751
remote sensing	250	**85.84**	**0.65**	**.824**	85.82	.69	.824	85.36	1.00	.818	84.72	1.17	.810	82.48	1.49	.783
	350	**86.51**	**0.39**	**.832**	86.14	.57	.827	85.83	0.51	.824	85.84	0.71	.824	83.38	1.38	.794
	450	**86.62**	**0.31**	**.833**	86.29	.37	.829	86.10	0.39	.827	86.39	0.48	.830	84.40	1.22	.806

the quality of the initial training samples. To shows the validity of the above statement, in the third experiment, we started the active learning process with biased initial training samples. To this end, for the ionosphere, the vowel and the remote sensing data sets, the initial training sets were defined by taking only 4, 12 and 16 labeled samples (two samples for each class), respectively, which are not sufficient to model the actual decision boundary of the classifier. Figs. 3(a), (b), and (c) show the average classification accuracies versus the number of samples included in the training set at each iteration obtained by different methods for the ionosphere, the vowel and the remote sensing data sets, respectively. From these figures, one can see that the proposed technique always provided higher classification accuracies compared to the other methods at the initial stage of learning process. From a different perspective, it can achieve the same accuracy of the other techniques with a significantly smaller number of samples. This confirms the robustness of the proposed technique to biased (poor) initial training samples.

The fourth experiment deals with the computational time required by the different techniques using the same experimental setting as described in the second experiment. All the experiments were carried out on a PC (INTEL(R) Core(TM)2 Duo 2.0 GHz with 2.0 GB of RAM). Table 4 shows the computational time (in seconds) required by the investigated techniques for all three data sets. From this table, one can see that the computational time required by the proposed approach is almost similar to the computational time taken by the MS-ABD approach. On the contrary, the computational time taken

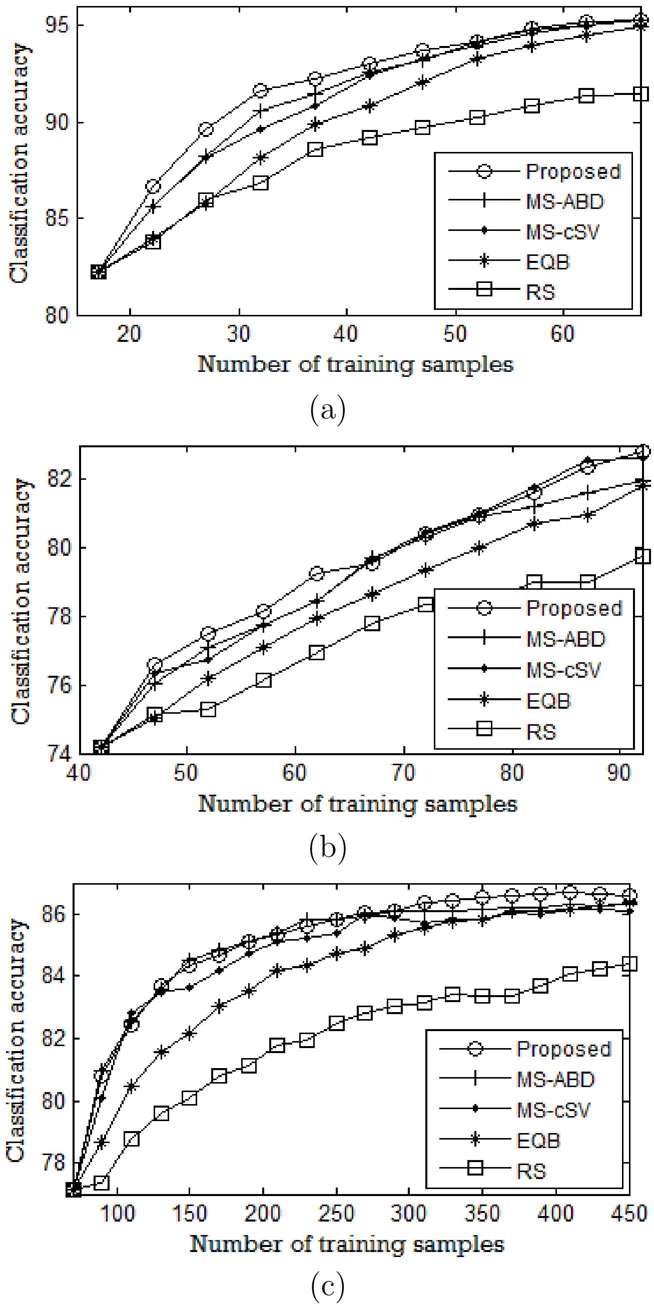
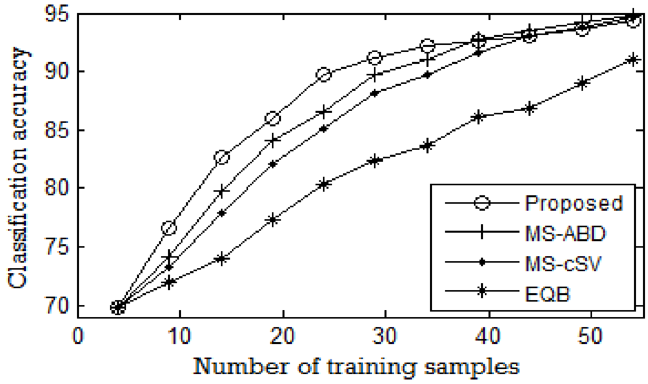


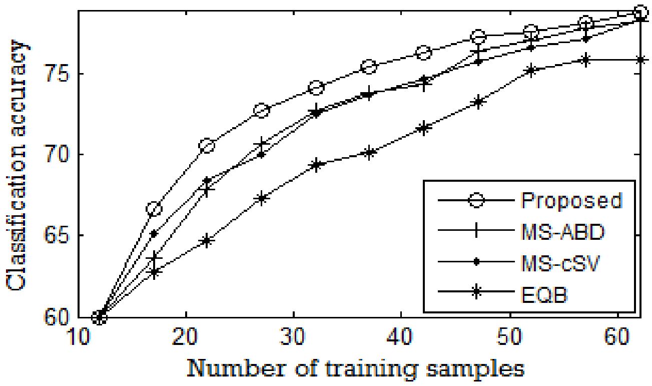
Fig. 2. Average classification accuracy over twenty runs versus the number of training samples provided by the Proposed, the MS-ABD, the MS-cSV, the EQB, and the RS methods for (a) the ionosphere, (b) the vowel, and (c) the remote sensing data sets.

by the MS-cSV and the EQB techniques is higher compared to the proposed technique. The RS method was obviously the most efficient in terms of computational load. Nonetheless, it resulted in the lowest classification accuracy.

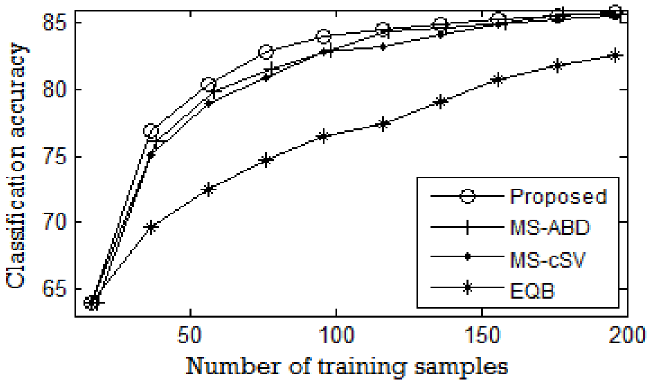
Finally, we carried out different experiments for assessing the stability of the proposed technique by varying the values of m and the width of the histogram



(a)



(b)



(c)

Fig. 3. Average classification accuracy provided by the Proposed, the MS-ABD, the MS-cSV, and the EQB methods for (a) the ionosphere, (b) the vowel, and (c) the remote sensing data sets by starting with biased training samples.

bins. The results of all these experiments pointed out the almost insensitivity of the proposed algorithm to these parameters.

Table 4
Computational time (in seconds) taken by the investigated active learning methods on the three real data sets considered

Methods	Data sets		
	ionosphere	vowel	remote sensing
Proposed	0.80	2.00	36.87
MS-ABD	0.80	1.97	34.66
MS-cSV	1.79	5.37	262.84
EQB	2.56	6.74	148.43
RS	0.48	1.22	14.32

6 Discussion and Conclusion

In this paper we have presented a novel active learning technique for solving multiclass classification problems with SVM classifiers by applying cluster assumption based uncertainty and kernel k-means based diversity criteria. To empirically assess the effectiveness of the proposed method we compared it with other three batch mode active learning techniques existing in the literature using a toy data set and three real data sets. In this comparison we observed that the proposed method always provided better (or comparable) accuracies with improved stability with respect to those achieved by some of the most effective techniques presented in the literature (i.e., the MS-ABD, the MS-cSV and the EQB). Moreover, it proved robust to handle biased initial training samples. Thus, in our experiments, the proposed algorithm provided the best trade-off among robustness to biased initial training samples, classification accuracy, computational complexity, and number of new labeled samples necessary to reach the convergence.

Acknowledgments

This work was carried out in the framework of the India-Trento Program for Advanced Research.

References

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

[2] S. Tong, D. Koller, Support vector machine active learning with applications to text classification, J. Machine Learning Research 2 (1) (2002) 45–66.

[3] P. Mitra, C. A. Murthy, S. K. Pal, A probabilistic active support vector learning algorithm, IEEE Trans. Pattern Analysis and Machine Intelligence 26 (3) (2004) 413–418.

[4] M. Li, I. K. Sethi, Confidence-based active learning, IEEE Trans. Pattern Analysis and Machine Intelligence 28 (8) (2006) 1251–1261.

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

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

[7] D. Tuia, F. Ratle, F. Pacifici, M. F. Kanevski, W. J. Emery, Active learning methods for remote sensing image classification, IEEE Trans. Geoscience and Remote Sensing 47 (7) (2009) 2218–2232.

[8] P. Rigollet, Generalization error bounds in semi-supervised classification under the cluster assumption, J. Machine Learning Research 8 (2007) 1369–1392.

[9] K. Fukumizu, Statistical active learning in multilayer perceptrons, IEEE Trans. Neural Networks 11 (1) (2000) 17–26.

[10] N. Roy, A. McCallum, Toward optimal active learning through sampling estimation of error reduction, in: Proc. 18th ICML, 2001, pp. 441–448.

[11] I. Dagan, S. P. Engelson, Committee-based sampling for training probabilistic classifiers, in: Proc. 12th ICML, 1995, pp. 150–157.

[12] Y. Freund, H. S. Seung, E. Shamir, N. Tishby, Selective sampling using the query by committee algorithm, Machine Learning 28 (1997) 133–168.

[13] S. Patra, L. Bruzzone, A fast cluster-assumption based active learning technique for classification of remote sensing images, IEEE Trans. Geoscience and Remote Sensing 49 (5) (2011) 1617–1626.

[14] B. Demir, C. Persello, L. Bruzzone, Batch-mode active-learning methods for the interactive classification of remote sensing images, IEEE Trans. Geoscience and Remote Sensing 49 (3) (2011) 1014–1031.

[15] C. J. C. Burges, A tutorial on support vector machines for pattern recognition, Data Mining and Knowledge Discovery 2 (2) (1998) 121–167.

[16] F. Melgani, L. Bruzzone, Classification of hyperspectral remote sensing images with support vector machines, IEEE Trans. Geoscience and Remote Sensing 42 (8) (2004) 1778–1790.

[17] M. Sezgin, B. Sankur, Survey over image thresholding techniques and quantitative performance evaluation, J. Electronic Imaging 13 (1) (2004) 146–165.

[18] J. N. Kapur, P. K. Sahoo, A. K. C. Wong, A new method for gray-level picture thresholding using the entropy of the histogram, Computer Vision, Graphics, Image Processing 29 (3) (1985) 273–285.

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

[20] A. Frank, A. Asuncion, UCI Machine Learning Repository, university of California, School of Information and Computer Sciences, `http://archive.ics.uci.edu/ml` (2010).

[21] S. K. Pal, S. Mitra, Multilayer perceptron, fuzzy sets, and classification, IEEE Trans. Neural Networks 3 (5) (1992) 683–697.

[22] C.-C. Chang, C.-J. Lin, LIBSVM: a library for support vector machine, software available at `http://www.csie.ntu.edu.tw/˜cjlin/libsvm` (2001).