Title: A Novel Classification Technique Based on Progressive Transductive SVM Learning

This paper appear in: Pattern Recognition Letters

Date of Publication: 2014

Author(s): Anshu Singla, Swarnajyoti Patra and Lorenzo Bruzzone

Volume: 42

Page(s): 101-106

DOI: 10.1016/j.patrec.2014.02.003

A Novel Classification Technique Based on Progressive Transductive SVM Learning

Anshu Singla [a], Swarnajyoti Patra [b,*], Lorenzo Bruzzone [c]

[a] *School of Mathematics and Computer Applications, Thapar University, Patiala 147004, India*

[b] *Department of Computer Science and Engineering, Tezpur University, Tezpur 784028, India*

[c] *Department of Information Engineering and Computer Science, University of Trento, Trento 38123, Italy*

Abstract

The existing semisupervised techniques based on progressive transductive support vector machine (PTSVM) iteratively select transductive samples that are closest to the SVM margin bounds. This may result in selecting wrong patterns (i.e., patterns that when included in the semisupervised learning can be associated with a wrong label) as transductive samples, especially when poor initial training sets are available or when available training samples are biased. To mitigate this problem, the proposed approach considers the distance from SVM margin bounds, the properties of the k-nearest neighbors approach, and the cluster assumption in the kernel space. To assess the effectiveness of the proposed method, we compared it with other PTSVM methods existing in the literature by using a toy data set and six real data sets. Experimental results confirmed the effectiveness of the proposed technique.

Key words: Cluster assumption, k-nearest neighbors, semisupervised classification, support vector machine, transductive inference.

1 Introduction

Supervised learning needs only labeled data for training. The classification results rely on the quantity and quality of these labeled samples. However the generation of proper labeled samples is often difficult, expensive and time consuming, as this requires the effort of experienced human annotators. On

* Corresponding author. email:swpatra@gmail.com

the other hand, unlabeled data are relatively easy to collect, although they have no use in supervised learning. Two popular machine learning approaches for dealing with this problem are active learning and semisupervised learning. Active learning expands the original training set according to an interactive process that involves a supervisor who can assign the correct label to unknown data points [1,2]. The goal of the active learner is to select the most informative samples so as to accurately learn from the fewest such additionally labeled data. In contrast, semisupervised learning exploits unlabeled data, together with the labeled data, to build better classifiers [3]. As a result, under proper assumptions, it requires less human effort for labeling. Thus, it becomes of great interest both in theory and in practice.

The transductive support vector machine (TSVM) is a semisupervised version of SVM [4]. During the training phase, it gradually (iterative process) searches a reliable separating hyperplane in the kernel space by taking into account both labeled and unlabeled samples. The unlabeled data can be used as an additional source of information about margin for SVM. In transductive learning the goal is to find a labeling of the unlabeled data, so that a linear boundary has the maximum margin on both the original labeled data and the (now labeled) unlabeled data (defined here as transductive data for convenience). This task can improve the generalization performance of SVMs, especially when poor training sets are available or when the available training samples are inadequate [5].

In the literature many semisupervised techniques based on TSVM exist for pattern classification [5–10]. All these methods try to find out a decision hyperplane passing through low density region of the kernel space. A good review of semisupervised approaches can be found in [11]. In [5], Joachims solved the quadratic optimization problem for the implementation of the TSVM with an application to text classification. This algorithm is effective when the ratio between the unlabeled positive and negative samples is known at the beginning of transductive learning. Chapelle and Zien [6] proposed a method that optimizes the transductive SVM objective function by using gradient descent technique to find out the decision boundary in low density regions of the kernel space. Sindhwani and Keerthi [7] proposed a fast algorithm for linear TSVM, suitable for large scale text applications. In [8], an additional criterion is included with the standard objective function of the TSVM and then a genetic algorithm is used for optimizing this objective function. In [9], a progressive TSVM algorithm was proposed that iteratively selects a positive and a negative sample as transductive samples from the available unlabeled samples that are inside the SVM decision margin and have minimum distance from positive margin and negative margin, respectively. The method also uses dynamic adjustment to reduce the miss labeling of selected transductive samples. In [10], Bruzzone et. al. modify the algorithm presented in [9] to select a batch of positive and negative patterns as transductive samples from the available

unlabeled patterns at each iteration of the transductive learning process.

The progressive TSVM (PTSVM) algorithms existing in the literature [9,10] iteratively select the most certain patterns as transductive samples from the available unlabeled patterns that are inside the SVM margin bounds. The certainty of a sample is measured by considering only its distance from the nearest SVM margin bound. This may result in a high probability of selecting wrong patterns as transductive samples (the actual labels of the patterns are different from the labels automatically assigned to them), especially when the initial decision hyperplane is poor i.e., passes through a wrong region of the kernel space. Thus, the final classification accuracy may be degraded. In this paper we propose a novel semisupervised technique based on PTSVM learning that mitigates the above-mentioned limitation. The proposed technique not only uses the distance from the nearest SVM margin bound but it also exploits the properties of k-nearest neighbors (k-nn) approach and the cluster assumption to selects the most certain samples as transductive samples at each iteration of the learning process.

The rest of this paper is organized as follows. The concept of inductive and transductive SVM learning is presented in Section 2. Section 3 describes the limitations of the existing PTSVM based approaches. The proposed technique is presented in Section 4. Section 5 provides the detailed description of the data sets used in the experiment and the results obtained on the considered data sets. Finally, Section 6 draws the conclusion of this work.

2 Support vector machine classifier

Before presenting the proposed semisupervised technique based on PTSVM learning, we briefly recall the main concepts associated with both the inductive SVM and transductive SVM learning. The reader may refer to [4,12] for more details on the SVM approach.

2.1 Inductive SVM learning

Let $X = \{(x_1, ..., x_n)|x_i \in \Re^d\}$ be the set of n available training samples and $Y = \{(y_1, ..., y_n)|y_i \in \{-1, +1\}\}$ be the set of associated labels. The standard SVM learning also called inductive SVM (ISVM) learning, tries to separate the data in the input space with the available training data by defining a hyperplane:

$$f(x) : wx + b = 0 \tag{1}$$

such that the distance between the closest vectors to the hyperplane is maximum. The maximal geometrical margin generated by the hyperplane is

$$\varphi(w) = \frac{2}{\parallel w \parallel} \tag{2}$$

In the case of linearly non-separable training data, the objective function of the ISVM learning is to find out a hyperplane by solving the following quadratic optimization problem

$$\begin{aligned} & \max_{\alpha} \left\{ \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} y_i y_j \alpha_i \alpha_j K(\mathbf{x}_i, \mathbf{x}_j) \right\} \\ \text{subject to: } & \sum_{i=1}^{n} y_i \alpha_i = 0 \\ & 0 \leq \alpha_i \leq C \end{aligned} \tag{3}$$

where ξ_i and α_i represents slack variables and Lagrangian multipliers, respectively. 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 [2].

2.2 Transductive SVM learning

In the above framework, to address the issue of the available biased/poor training set, a transductive SVM approach has been proposed by involving unlabeled samples into the training phase [4]. Let $X^* = \{(x_1^*, ..., x_m^*) | x_i \in \Re^d\}$ be the set of m unlabeled samples and $Y^* = \{(y_1^*, ..., y_m^*) | y_i \in \{-1, +1\}\}$ be the corresponding predicted labels. At the initial iteration, the standard ISVM is used to obtain a separating hyperplane using the training set X only. Then, depending on the distance from this hyperplane labels are assigned to the unlabeled samples which are thus called semilabeled data. After that, according to a defined criterion transductive samples chosen from the semilabeled patterns are included into the original training set X. The resulting training set is used at the following iterations to find a more reliable discriminant hyperplane. This hyperplane separates (X, Y) and (X^*, Y^*) with the maximal margin and is derived as follows:

$$\varphi(w) = \min_{w,\xi_i,\xi_u^*} \left\{ \frac{1}{2} \| w \|^2 + C \sum_{i=1}^{n} \xi_i + C^* \sum_{i=1}^{d} \xi_u^* \right\}$$
$$\text{subject to: } \forall_{i=1}^{n} : y_i(wx_i + b) \geq 1 - \xi_i, \xi_i > 0$$
$$\forall_{i=1}^{m} : y_i^*(wx_i + b) \geq 1 - \xi_u^*, \xi_u^* > 0 \quad (4)$$

In order to handle the nonseparable training and transductive data, similarly to the ISVMs, the slack variables ξ_i and ξ_u^* and the associated penalty values C and C^* of both the training and transductive samples are introduced. In the learning process of the TSVMs, the purpose of C and C^* is to control the number of misclassified samples that belong to the original training set and to the unlabeled set, respectively. On increasing their values, the penalty associated with the errors on the training and transductive samples increases. In other words, the larger the regularization parameter are, the higher is the influence of the associated samples on the selection of the discriminant hyperplane. $d(d \leq m)$ is the number of transductive samples chosen at each iteration of transductive learning.

3 Limitations of the existing methods

The semisupervised techniques based on PTSVM existing in the literature [9,10], iteratively select transductive samples from semilabeled patterns that are inside and closest to the SVM margin bounds. This results in a high probability of selecting wrong patterns as transductive samples when the initial SVM decision hyperplane is very poor i.e., in cases where it passes through wrong region of the kernel space. Fig. 1 (a) depicts a situation where one SVM margin bound is near to the actual decision hyperplane (i.e., it passes through low density region of the kernel feature space) and the other one is far from it. Now, if we apply the conventional PTSVM methods to select transductive samples, they will select some semilabeled patterns as transductive samples that also shift the margin bound which passes through low density region towards wrong direction. As a result a poor decision hyperplane will be generated. The main reason is that these algorithms do not impose any explicit constraint to the transductive sample selection process so that the margin bound which passes through high density region can only be shifted towards low density region. Fig. 1 (b) depicts another extreme situation where both positive and negative margin bounds pass through both the classes. Also in this case the conventional PTSVM methods may fail to select the appropriate transductive samples. As a result, the classification performance can be degraded. In this paper we propose a novel semisupervised technique based on PTSVM learning that addresses both the above-mentioned issues during the selection of trasductive samples at each iteration of the learning process.

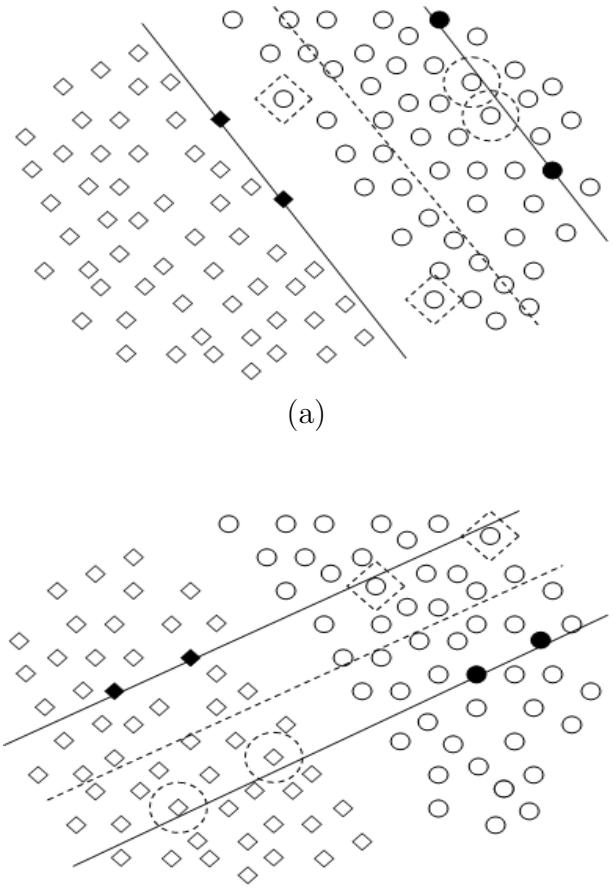

Fig. 1. Patterns belonging to class "-1" and "+1" are shown as white squares and circles, respectively. Initial labeled samples for class "-1" and "+1" are shown as black squares and circles, respectively. The separation hyperplane is shown as a dashed line, whereas the solid lines define the margin. The dashed squares and circles highlight the transductive patterns selected by the 1st iteration of conventional progressive TSVM methods labeled as "-1" and "+1", respectively (a) when one decision margin is near to the actual decision hyperplane and the other one is far from it; (b) when both positive and negative decision margins pass through the feature space of both the classes.

4 Proposed TSVM method

As regards the selection of transductive patterns, two points should be considered: 1) choose the informative samples, and 2) select the samples with an expected accurate labeling. In the proposed work the informative samples are selected by considering only the unlabeled patterns inside the SVM margin bounds. To select the proper transductive samples (with expected accurate labeling), in this paper we not only considered the distance from the SVM

margin bounds, yet we also exploits k-nn technique and the cluster assumption. The details of the proposed transductive samples selection procedure is given below.

Let us consider at the beginning only few training samples $X = \{(x_1, ..., x_n)|x_i \in \Re^d\}$; with their associated labels $Y = \{(y_1, ..., y_n)|y_i \in (-1, +1)\}$. Let us assume that a large number of unlabeled samples $X^* = \{(x_1^*, ..., x_m^*)|x_i \in \Re^d\}$ are available. First, we train the SVM classifier using the training set X only (inductive learning) to find out the decision hyperplane $f(x)$. Then, depending on the distance from this hyperplane, pseudo labels are assigned to all the unlabeled samples, thus obtaining semilabeled samples. Let N be an integer variable. We fix the value of $N = min\{N_{SV}^+, N_{SV}^-\}$, where N_{SV}^+ and N_{SV}^- represent the number of positive and negative margin support vectors, respectively, obtained after the first inductive learning iteration. After that, the set ψ^+ of positive candidate transductive samples is initialized by choosing N semilabeled samples from X^* that lie inside the margin bound and are closest to the positive margin. Similarly the negative candidate set ψ^- is initialized by choosing N semilabeled samples from X^* that lie inside the margin bound and are closest to the negative margin. If we select all the semilabeled samples in ψ^+ and ψ^- as transductive samples as in the existing techniques then the problem shown in Fig. 1 (a) may arises (i.e., the decision hyperplane may shift towards wrong direction). To mitigate this kind of problem, here first we compute the average distance d^+ and d^- of all samples in ψ^+ and ψ^- from the SVM decision hyperplane as follows:

$$d^+ = \frac{\sum_{x_i \in \psi^+} f(x_i)}{N} \tag{5}$$

$$d^- = \frac{\sum_{x_i \in \psi^-} |f(x_i)|}{N} \tag{6}$$

Then a threshold value t is obtained as $t = min\{d^+, d^-\}$. After finding the value of threshold t we update the positive candidate set $\psi^+ = \{x_i|x_i \in \psi^+; f(x_i) > t\}$ and negative candidate set $\psi^- = \{x_i|x_i \in \psi^-; |f(x_i)| > t\}$. This help us to incorporate cluster assumption criterion to choose appropriate transductive samples.

Due to the second problem shown in Fig. 1(b), there may be some semilabeled patterns that are closest to the positive (or negative) margin but belongs to ψ^- (or ψ^+). To select appropriate transductive samples from $\psi^\pm$, in this work we propose to apply a k-nn technique in the SVM kernel space. For each pattern $x_i^* \in \psi^\pm$, we find out the k-nearest labeled samples from X and then assign a class label to x_i^* according to k-nearest neighbor rule. After that according to the k-nn rule, the patterns from ψ^+ and ψ^- selected as transductive patterns those have positive and negative class label, respectively. To compute the

distance between a pattern $x_i^* \in \psi^{\pm}$ and $x_j \in X$ in the kernel space we use the same kernel function as used in the SVM classifier. Accordingly we try to select transductive samples that are most certain. The certainty of each unlabeled samples is measured by applying cluster assumption criterion and by exploiting the properties of the SVM and the k-nn techniques. Note that compared to the existing methods, the proposed technique spends some additional time to find out the k-nearest neighbors of each sample in $\psi^{\pm}$. Since a limited number of samples are in $\psi^{\pm}$, this does not take significant additional time.

Let ψ_t be the set of transductive samples. Initially the set is empty. After selecting few semilabeled samples as transductive samples we add them into the transductive set ψ_t and retrain the SVM classifier using the available labeled set X and ψ_t. The process is iterated until some patterns are inside the margin bound or a threshold value on the number of patterns in the margin is reached. If the label of a transductive pattern at iteration $itr+1$ is different from the one at iteration itr, such pattern is removed from the transductive set and reassigned to the unlabeled set. The regularization parameter for the transductive patterns C^* is increased in a linear way, depending on the value of the regularization parameter for the labeled patterns C and a parameter g. In our algorithm, the initial value of C^* is defined as $C^* = g \times C$, where g is a weight parameter computed as $g = \frac{1}{l} \times itr$ (l is the growth rate and is a user defined parameter). In the first l iterations, the value of g increases linearly and reach value equal to 1 at l^{th} iteration. After that the value of g does not change by increasing the iteration number.

It is worth noting that to solve multiclass problem we adopt one-against-all binary SVM architecture. Thus, if there are n classes then n TSVM are defined. Each TSVM solves a binary classification problem defined by one information class against all the others. The details of the proposed technique for solving multiclass problem is shown in algorithm 1.

Algorithm 1: Proposed transductive learning technique

Let X and X^* denote the available labeled and unlabeled samples and n represent the number of classes.

For s=1 to n

Step 1: Set $itr = 0$ and transductive candidate set $\psi_t = \phi$.

Repeat

Step 2: Train the s^{th} OAA binary SVM with the available labeled samples X to generate the decision function $f_s(.)$.

Step 3: Fix $N = \min(N_{SV}^+, N_{SV}^-)$.

Step 4: Define the positive candidate set ψ^+ and negative candidate set ψ^- by selecting N samples from X^* that lies inside the margin bound and are closest to the positive and negative margins, respectively.

Step 5: Compute average distances $d^{\pm}$ and threshold t as follows:
$d^+ = \frac{\sum_{x_i \in \psi^+} f(x_i)}{N}$, $d^- = \frac{\sum_{x_i \in \psi^-} |f(x_i)|)}{N}$ and $t = min\{d^+, d^-\}$.

Step 6: Update
$\psi^+ = \{x_i^* \mid x_i^* \in \psi^+, |f_s(x_i^*)| \geq t\}$; $\psi^- = \{x_i^* \mid x_i^* \in \psi^-, |f_s(x_i^*)| \geq t\}$

Step 7: For each $x_i^* \in \psi^{\pm}$, compute the distance to all the samples $x_j \in X$ in SVM kernel space. Then according to k-nearest neighbor rule assign either positive or negative class label to each $x_i^* \in \psi^{\pm}$.

Step 8: Update
$\psi^+ = \{x_i^* \mid x_i^* \in \psi^+$ and positive label assigned by $k - nn$ rule$\}$;
$\psi^- = \{x_i^* \mid x_i^* \in \psi^-$ and negative label assigned by $k - nn$ rule$\}$

Step 9: Update $\psi_t = \{\psi^+ \cup \psi^-\}$ and $X^* = X^* - \psi_t$

Step 10: $itr = itr + 1$;

Step 11: If $itr \leq l$ then update $g = \frac{1}{l} \times itr$ and $C^* = C \times g$.

Step 12: Train s^{th} binary SVM with updated labeled set $(X \cup \psi_t)$.

Step 13: For a sample $x_i^* \in \psi_t$, if sign of $f_s^{itr}(x_i^*)$ and $f_s^{itr-1}(x_i^*)$ is different then update $\psi_t = \psi_t - x_i^*$ and $X^* = X^* \cup x_i^*$.

Until $x_i^* \in X^*$ is found inside the margin bound or a threshold value on the number of patterns in the margin is reached.

End For

5 Experimental results

5.1 Description of data sets

In order to assess the effectiveness of the proposed technique, seven data sets with significantly different properties were used in the experiment. The first one is a toy data set which is made up of four linearly separable classes as shown in Fig. 2. The second one is a more complicated vowel data set [13]. The rest five data sets (Iris, Diabetes, Letter Recognition, Ionosphere and ISOLET) are taken from UCI machine learning repository among those that

are widely used as benchmark for pattern classification tasks[14]. For all data sets, first only few available labeled samples were randomly selected as initial training set X, and the rest were stored in the unlabeled pool X^*. Table 1 shows the details of all the above mentioned data sets.

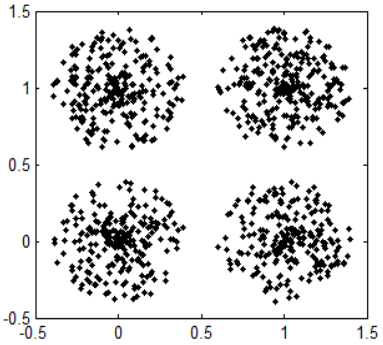

Fig. 2. Linearly separable toy data set.

Table 1
Number of patterns, features and classes for the seven different data sets

Data sets	**Patterns**	**Features**	**Classes**
Toy	1000	2	4
Vowel	871	3	6
Iris	150	4	3
Diabetes	768	8	2
Letter Recognition	20000	16	26
Ionosphere	351	34	2
ISOLET	6238	617	26

5.2 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 [15]. 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-nn technique. The value of k for kernel k-nn is also automatically computed using cross-validation technique. To gradually consider the influence of transductive samples for defining the actual decision hyperplane, the initial value of the regularization parameter for the selected transductive patterns C^* should be small since at initial stage of learning the SVM decision hyperplane is poorly defined. At subsequent iterations, since

labeling confidence increases, the value of C^* will also increase. This is done by defining the value of the growth rate parameter l. In our experiments, the value of growth rate l is fixed to 10 for all the considered data sets.

To assess the effectiveness of the proposed technique we compared it with two other existing methods: i) the progressive transductive SVM (PTSVM) method [9]; and ii) the modified progressive transductive SVM (MPTSVM) method [10]. In PTSVM, at each iteration of transductive learning, a positive and a negative semilabeled pattern are selected as transductive samples from the available samples that are inside the SVM margin bound and are closest to the positive and negative margins, respectively. In MPTSVM, at each iteration, a batch of positive and negative semilabeled patterns are selected as transductive samples using the same criterion as used in PTSVM. The batch size is determined automatically depending on the number of available positive and negative support vectors at a particular iteration of the learning process.

The multiclass SVM with the standard OAA architecture has been manually implemented by using the LIBSVM library (for Matlab interface) [16]. All the algorithms presented in this paper have been implemented in Matlab.

5.3 Results

In order to understand the effectiveness of the proposed technique, in the first experiment we compared the performance of the proposed method with the existing PTSVM and MPTSVM methods. For all the four data sets, initially only few labeled samples were considered in the training set. The transductive learning process was repeated for 10 trials with 10 different initial training sets (generated randomly) to reduce the random effect on the results. Table 2 shows the average overall classification accuracies and standard deviations provided by different methods starting with different numbers of initial labeled samples for the toy, the iris, the diabetes and the vowel data sets. From the table one can see that for the toy data set, the proposed technique yielded a classification accuracy of 100% starting with only 8 initial labeled samples, whereas both the PTSVM (95.90%) and the MPTSVM (97.10%) techniques failed to achieve the same accuracy under the same conditions. For the vowel, the iris, the diabetes and the letter recognition data sets, the proposed technique starting with different numbers of initial labeled samples always resulted in higher classification accuracy than the other techniques. For example, considering the iris data set with 18 initial labeled samples, the proposed technique resulted in a 96.47% classification accuracy, whereas the best accuracy obtained by the existing literature methods is 94.87%. Similarly, for the vowel, the diabetes and the letter recognition data sets, one can see that the proposed technique always achieved at least 2% higher classification accuracies compared to those

obtained by the literature methods under the same conditions. Moreover, from Table 2 one can see that the standard deviation of the accuracy provided by the proposed approach with respect to the different trials 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. Since the proposed technique selects the transductive samples not only considering the distance from the SVM margin bounds, yet it also exploits the properties of the k-nn technique and the cluster assumption, it reduces the probability of selecting wrong transductive samples during the learning. Thus, it is more robust to solve classification tasks. For the ionosphere and the ISOLET data sets, the proposed technique only slightly improved the accuracy as compared to the literature techniques. This may be due to the fact that the cluster assumption criterion incorporated by the proposed technique may be less effective for these high dimensional data sets.

Table 2
Average overall classification accuracy (acc) and its standard deviation (std) obtained on ten runs starting with different number of initial labeled samples ($|X|$)

Data sets	$\|X\|$	**Proposed**		**PTSVM**		**MPTSVM**	
		acc	**std**	**acc**	**std**	**acc**	**std**
Toy	8	100	0.00	95.90	3.68	97.10	1.83
Vowel	42	76.04	2.42	74.06	2.66	74.08	3.16
	56	77.93	2.14	75.92	2.66	75.98	2.85
Iris	9	95.12	3.83	93.47	5.21	93.53	4.69
	18	96.47	3.41	94.87	4.15	94.87	4.27
Diabetes	11	69.42	3.62	65.47	5.94	66.99	5.44
	22	70.68	3.87	67.59	7.23	67.50	6.86
Letter Recognition	280	68.93	2.79	65.81	2.58	66.32	2.17
	561	77.32	1.21	71.74	1.45	73.37	1.82
Ionosphere	54	89.56	2.46	88.80	3.30	89.46	2.67
	69	92.02	1.51	91.33	1.38	91.34	1.18
ISOLET	208	82.21	2.50	81.53	2.73	81.72	2.28
	338	87.04	1.52	86.27	1.78	86.48	1.67

The second set of experiment was devoted to analyze the performance of the proposed technique by varying the value of the regularization parameter for transductive patterns C^*. As explained in Section 4, the initial and incremented (after completion of an iteration) value of C^* are computed based on the value of the user defined growth rate parameter l. Thus, in this experiment for all considered data sets l was varied in the range 5, 10, 15 and 20. Table

3 shows the classification accuracies obtained by the proposed technique for different values of l. From the table one can see that the classification accuracy is not significantly affected. Thus, the value of C^* can be computed by assigning the value of l in wide range.

Table 3
Average classification accuracy provided by the proposed approach considering different values of the user defined parameter l used to compute the value of C^*.

Data sets	$\|X\|$	**Proposed technique**			
		l=5	**l=10**	**l=15**	**l=20**
Toy	8	100	100	100	100
Vowel	56	77.71	77.93	78.12	77.83
Iris	18	96.42	96.47	96.58	96.51
Diabetes	22	70.75	70.68	70.62	70.57
Letter Recognition	561	77.27	77.32	77.05	76.96
Ionosphere	69	91.81	92.02	92.16	92.08
ISOLET	338	86.87	87.04	87.13	86.95

6 Discussion and conclusion

In this paper we have proposed a novel semisupervised technique based on PTSVM for solving pattern classification tasks, which overcomes the limitations of the existing PTSVM based methods. The existing techniques select the transductive samples by exploiting only the properties of the SVM classifier. They do not take adequately into account the low-density region of the feature space as well as the possible poor initial training set in the definition of the criterion for selecting transductive samples. As a result, the probability of selecting wrong transductive patterns (the actual labels of the patterns are different from the automatically assigned labels) becomes high and the classification performance may result degraded. To overcome this problem, the proposed technique not only exploits the properties of the SVM classifier, yet it also exploits a k-nn technique and the cluster assumption for selecting accurate transductive samples.

To empirically asses the effectiveness of the proposed method we compared it with other PTSVM based approaches existing in the literature using a toy data set and six real data sets. By this comparison we observed that the proposed method provided better accuracy compared to the existing techniques on the considered data sets.

Compared to the existing methods the proposed technique needs some additional time for using the k-nn technique. However, since the proposed technique computes the k-nearest neighbors of few samples this does not take significant additional time.

References

[1] S. Patra, L. Bruzzone, A batch-mode active learning technique based on multiple uncertainty for SVM classifier, IEEE Geoscience and Remote Sensing Letters 9 (3) (2012) 497–501.

[2] S. Patra, L. Bruzzone, A cluster-assumption based batch mode active learning technique, Pattern Recognition Letters 33 (9) (2012) 1042–1048.

[3] K. Nigam, A. McCallum, T. Mitchell, Learning to classify text from labeled and unlabeled documents, in: Proc. AAAI, 1998, pp. 792–799.

[4] V. N. Vapnik, Statistical Learning Theory, Willy, New York, 1998.

[5] T. Joachims, Transductive inference for text classification using support vector machines, in: Proc. 16th ICML, 1999, pp. 200–209.

[6] O. Chapelle, A. Zien, Semi-supervised classification by low density separation, in: Proc. 10th International Workshop on Artificial Intelligence and Statistics, 2005.

[7] V. Sindhwani, S. S. Keerthi, Large scale semisupervised linear SVMs, in: Proc. SIGIR, 2006.

[8] M. M. Adankon, M. Cheriet, Learning semi-supervised SVM with genetic algorithm, in: Proc. Int. conf. on Neural Networks, 2007.

[9] Y. Chen, G. Wang, S. Dong, Learning with progressive transductive support vector machine, Pattern Recognition Letters 24 (2003) 1845–1855.

[10] L. Bruzzone, M. Chi, M. Marconcini, A novel transductive SVM for semisupervised classification of remote-sensing images, IEEE Trans. Geoscience and Remote Sensing 44 (11) (2006) 3363–3373.

[11] X. Zhu, Semi-supervised learning literature survey, Tech. Rep. 1530, Computer Sciences, University of Wisconsin-Madison (2005).

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

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

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

[15] 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.

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