

# Support Vector Selection and Adaptation and Its Application in Remote Sensing

Gülşen Taşkın Kaya  
Computational Science and Engineering  
Istanbul Technical University  
Istanbul, Turkey  
[gtaskink@purdue.edu](mailto:gtaskink@purdue.edu)

Okan K. Ersoy  
School of Electrical and Computer Engineering  
Purdue University  
W. Lafayette, IN, USA  
[ersoy@purdue.edu](mailto:ersoy@purdue.edu)

Mustafa E. Kamaşak  
Computer Engineering  
Istanbul Technical University  
Istanbul, Turkey  
[kamasak@itu.edu.tr](mailto:kamasak@itu.edu.tr)

**Abstract**—Classification of nonlinearly separable data by nonlinear support vector machines is often a difficult task, especially due to the necessity of a choosing a convenient kernel type. Moreover, in order to get high classification accuracy with the nonlinear SVM, kernel parameters should be determined by using a cross validation algorithm before classification. However, this process is time consuming. In this study, we propose a new classification method that we name Support Vector Selection and Adaptation (SVSA). SVSA does not require any kernel selection and it is applicable to both linearly and nonlinearly separable data. The results show that the SVSA has promising performance that is competitive with the traditional linear and nonlinear SVM methods.

**Keywords**-Support Vector Machines; Classification of Remote Sensing Data; Support Vector Machines; Support Vector Selection and Adaptation.

## I. INTRODUCTION

Support Vector Machine is a machine learning algorithm, developed by Vladimir Vapnik, used for classification or regression [1]. This method can be used for classification of linearly and nonlinearly separable data. Linear SVM uses a linear kernel and nonlinear SVM uses a nonlinear kernel to map the data into a higher dimensional space in which the data can be linearly separable. For nonlinearly separable data, nonlinear SVM generally performs better compared to the linear SVM.

The performance of nonlinear SVM depends on the kernel selection [2]. It has been observed that apriori information about the data is required for the selection a kernel type. Without such information, choosing a kernel type may not be easy.

It is possible to try all types of kernels and to select the one that gives the highest accuracy. For each trial, kernel parameters have to be tuned for highest performance. Therefore, this is a time-consuming approach.

In order to overcome these difficulties, we have developed a new machine learning algorithm that we called Support Vector Selection and Adaptation (SVSA). This algorithm starts with the support vectors obtained by linear SVM. Some of these support vectors are selected as reference vectors to increase the classification performance. The algorithm is finalized by adapting the reference vectors with respect to training data [3]. Testing data are classified by using these reference vectors with K-Nearest neighbor method (KNN) [4].

During our preliminary tests with SVSA, we have observed that it outperforms the linear SVM, and it has close classification accuracy compared to the nonlinear SVM. The proposed algorithm is tested on both synthetic data and remote sensing images.

In this work, the performance of the proposed SVSA algorithm is compared to other SVM methods using two different datasets: Colorado dataset with 10 classes and 7

features and Panchromatic SPOT images recorded before and after earthquake, occurred on 17 August 1999 in Adapazari.

## II. SUPPORT VECTOR SELECTION AND ADAPTATION

The SVSA method starts with the support vectors obtained from linear SVM, and it eliminates some of them for not being sufficiently useful for classification. Finally, the selected support vectors are modified and used as reference vectors for classification. In this way, nonlinear classification is achieved without a kernel.

### A. Support Vector Selection

The SVSA has two steps: Selection and adaptation. In the selection step, the support vectors obtained by the linear SVM method are classified using KNN. Afterwards, the misclassified support vectors are removed from the set of support vectors, and the remaining vectors are selected as the reference vectors as a candidate for the adaptation process.

Let  $\underline{X} = \{(x_1, \bar{x}_1), \dots, (x_N, \bar{x}_N)\}$  represent the training data with  $x_i \in R^p$  and the class labels  $\bar{x}_i \in \{1, \dots, K, M\}$ .  $N$ ,  $M$  and  $p$  denote the number of training samples, the number of classes and the number of features respectively. After applying the linear SVM to the training data, the support vectors are obtained as

$$\underline{S} = \{(s_i, \bar{s}_i) \mid (s_i, \bar{s}_i) \in X \quad i = 1, \dots, K, k\} \quad (1)$$

$$\underline{T} = \{(t_i, \bar{t}_i) \mid (t_i, \bar{t}_i) \in X \setminus S \quad i = 1, \dots, K, N - k\} \quad (2)$$

where  $k$  is the number of support vectors,  $S$  is the set of support vectors with the class labels  $s$ , and  $T$  is the set of training data vectors with the class labels  $t$ , excluding the support vectors.

In the selection stage, the support vectors in the set  $S$  are classified with respect to the set  $T$  by using the KNN algorithm. The labels of the support vectors are obtained as:

$$\bar{s}_i^p = \left\{ \bar{t}_l \mid l = \arg \min_{1 \leq j \leq N-k} \|s_i - t_j\| \right\} \quad i = 1, \dots, K, k \quad (3)$$

where  $\bar{s}_i^p$  is the predicted label of the  $i^{th}$  support vector.

The misclassified support vectors are then removed from the set  $S$ . The remaining support vectors are called reference vectors and constitute the set  $\bar{R}$ :

$$\bar{R} = \left\{ (s_i, \bar{s}_i) \mid (s_i, \bar{s}_i) \in S \quad \text{and} \quad \bar{s}_i^p = \bar{s}_i \quad i = 1, \dots, K, k \right\} \quad (4)$$

The aim of the selection process is to select the support vectors which best describe the classes in the training set.

### B. Adaptation

In the adaptation step, the reference vectors are adapted with respect to the training data by moving them towards or away from the decision boundaries. The corresponding adaptation process used is similar to the Learning Vector Quantization (LVQ) algorithm described as below [5,6].

Let  $x_j$  be one of the training samples with label  $y_j$  [7]. Assume that  $\mathbf{r}_w(t)$  is the nearest reference vector to  $x_j$  with label  $y_{r_w}$ . If  $y_j \neq y_{r_w}$  then the adaptation is applied as follows:

$$\mathbf{r}_w(t+1) = \mathbf{r}_w(t) - \eta(t)(\mathbf{x}_j - \mathbf{r}_w(t)) \quad (5)$$

On the other hand, if  $\mathbf{r}_l(t)$  is the nearest reference vector to  $x_j$  with label  $y_{r_l}$  and  $y_j = y_{r_l}$  then

$$\mathbf{r}_l(t+1) = \mathbf{r}_l(t) + \eta(t)(\mathbf{x}_j - \mathbf{r}_l(t)) \quad (6)$$

where  $\eta(t)$  is a descending function of time called the learning rate. It is also adapted in time by

$$\eta(t) = \eta_0 e^{-t/\tau} \quad (7)$$

where  $\eta_0$  is the initial value of  $\eta$ , and  $\tau$  is a time constant.

At the end of the adaptation process, the reference vectors are used in the classification. 1-Nearest Neighbor classification with all the reference vectors is used to make a final classification. The aim of the adaptation process is to make the reference vectors distribute around the decision boundary of classes, especially if the data are not linearly separable.

## III. REMOTE SENSING APPLICATIONS

In order to compare the classification performance of our method with other SVM methods, two different remote sensing dataset are used.

TABLE I. TRAINING AND TESTING SAMPLES OF THE COLORADO DATASET

Class	Type of Class	#Testing Data	#Testing Data
1	Water	408	195
2	Colorado Blue Spruce	88	24
3	Mountane/ Subalpine meadow	45	42
4	Aspen	75	65

Class	Type of Class	#Testing Data	#Testing Data
5	Ponderosa Pine 1	105	139
6	Ponderosa Pine/Douglas Fir	126	188
7	Engelmann Spruce	224	70
8	Douglas Fir/White Fir	32	44
9	Douglas Fir/Ponderosa Pine/Aspen	25	25
10	Douglas Fir/White Fir/Aspen	60	39
<b>Total</b>		<b>1188</b>	<b>831</b>

TABLE II. TRAINING CLASSIFICATION ACCURICIES FOR THE COLORADO DATASET

Methods	Classification Performance Percentage of Training Data										
	Class 1	Class 2	Class 3	Class 4	Class 5	Class 6	Class 7	Class 8	Class 9	Class 10	Overall
SVM	100.00	67.05	51.11	53.33	8.57	87.30	90.18	37.50	0.00	45.00	74.92
NSVM(1)	100.00	100.00	55.56	86.67	42.86	84.92	98.66	53.13	64.00	71.67	87.12
NSVM(2)	100.00	73.86	33.33	37.33	0.00	78.57	89.29	0.00	0.00	0.00	68.60
SVSA	100.00	100.00	75.56	90.67	93.33	84.92	97.32	87.50	72.00	85.00	<b>94.11</b>

TABLE III. TESTING CLASSIFICATION ACCURICIES FOR THE COLORADO DATASET

Methods	Classification Performance Percentage of Testing Data										
	Class 1	Class 2	Class 3	Class 4	Class 5	Class 6	Class 7	Class 8	Class 9	Class 10	Overall
SVM	100.00	37.50	4.76	33.85	3.60	59.04	92.86	0.00	0.00	20.51	50.18
NSVM(1)	94.36	91.67	2.38	36.92	1.44	47.34	100.00	0.00	0.00	69.23	50.42

Since the first dataset has ten classes, the SVSA algorithm is generalized to a multi-class algorithm by using one-against-one approach [8].

Moreover, all the data are scaled to decrease the range of the features and to avoid numerical difficulties during the classification. For nonlinear SVM method, the kernel parameters are determined by using ten fold cross-validation [9].

#### A. The Colorado Dataset

Classification is performed with the Colorado dataset consisting of the following four data sources [10]:

- Landsat MSS data (four spectral data channels),
- Elevation data (one data channel),
- Slope data (one data channel),
- Aspect data (one data channel).

Each channel comprised an image of 135 rows and 131 columns, and all channels are spatially co-registered. It has ten

ground-cover classes listed in Table 1. One class is water; the others are forest types. It is very difficult to distinguish among the forest types using Landsat MSS data alone since the forest classes show very similar spectral response.

All these classes are classified by multiclass SVSA, linear SVM and nonlinear SVM with radial basis and polynomial kernel, respectively. The classification accuracy for each class and overall classification accuracies of the methods are listed in Table 2.

According to the results in Table 2, the overall classification performance is generally quite low for all methods since the Colorado dataset is a difficult classification problem. The overall classification accuracy of the SVSA is better than the other methods. In addition, it gives high classification accuracy for many classes individually in comparison to nonlinear SVM.

#### B. SPOT HRVIR Images in Adapazari, Turkey

SPOT HRVIR Panchromatic images were captured on 25 July 1999 and 4 October 1999 with a spatial resolution of 10

meters. They were geometrically corrected using 26 ground control points from 1:25 000 topographic maps of the area. Images were transformed to Universal Transverse Mercator (UTM) coordinates using a first order polynomial transformation and nearest neighbor re-sampling [11].

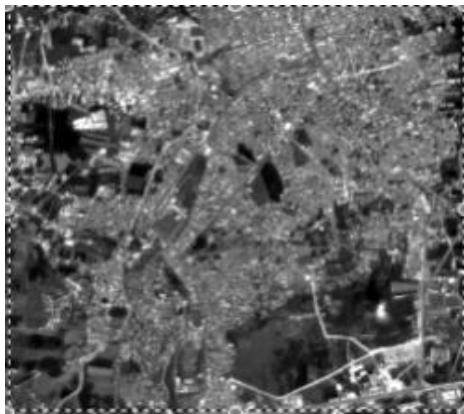


Figure 1: Panchromatic image captured in 25 July 1999 (some region of pre-earthquake image in Adapazari).

Initially, the urban and vegetation area are classified by using the intensity values of pre-earthquake image with the SVSA method, and then a thematic map is created with two classes (Figure 2): Urban area and vegetation area.

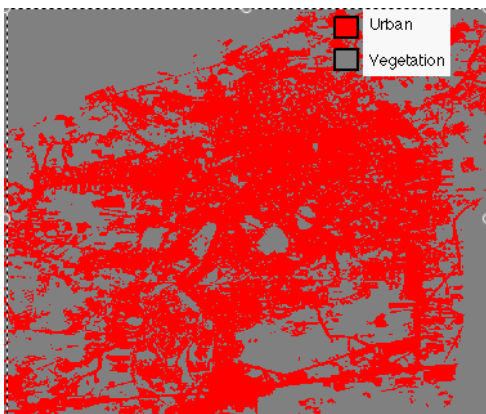


Figure 2: Classified thematic map obtained by applying the SVSA method to pre-earthquake image.

- In the second step, the SVSA method was applied to difference image obtained from the subtraction of post and pre image matrix. However, in this case, the method is applied to only urban regions within the difference image with two classes; collapsed and uncollapsed buildings.

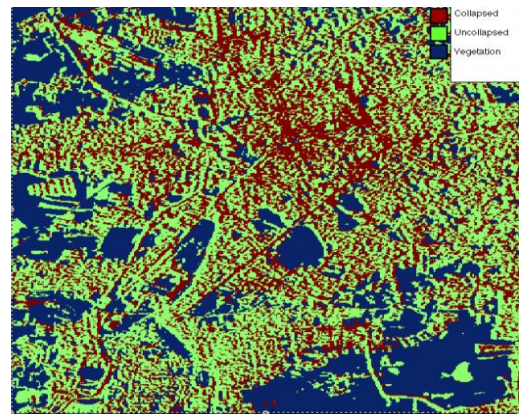


Figure 3: Collapsed buildings indicated by the SVSA from difference image.

Vegetation regions may change during time. Therefore, the vegetation areas can be misinterpreted as collapsed buildings. In order to avoid this, the SVSA method is applied only to the urban regions.

Since the SVSA method is a supervised learning algorithm like the other SVM methods as well, it requires having a training dataset with their label information for all the classes to be classified. Because of that, the training dataset for urban and vegetation area were taken from the pre-earthquake image. The training data for collapsed buildings were taken from the difference image because it is easier to visually pick the collapsed samples.

The pre-earthquake images are used to classify the urban and vegetation areas. Afterwards, ten different combinations of these dataset are randomly created, and all the methods are applied for each dataset individually.

Box plots of Macro-F error rates on these dataset summarize the average F scores on the two classes in Figure 4. Our algorithm has very low error rates and very small deviations compared to linear and nonlinear SVM with polynomial kernel (NSVM(2)). In addition, the SVSA method

has competitive classification performance compared to nonlinear SVM with radial basis kernel (NSVM(1)).

the SVSA results and make a consensus between these two methods for linear data.

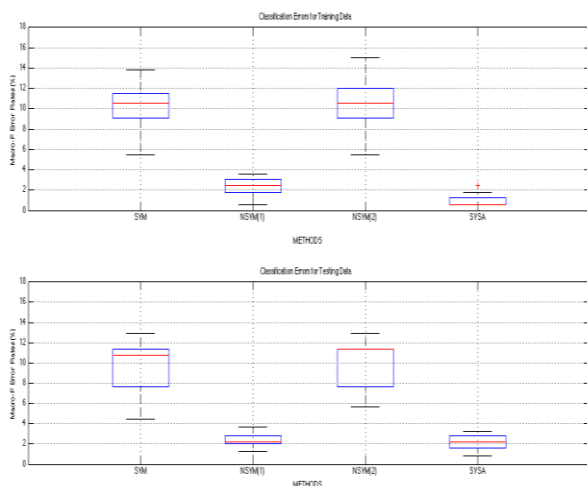


Figure 4: Collapsed buildings indicated by the SVSA from difference image.

#### IV. CONCLUSION

In this study, we addressed the problem of classification of remote sensing data using the proposed support vector selection and adaptation (SVSA) method in comparison to linear and nonlinear SVM.

The SVSA method consists of selection of the support vectors that contribute most to the classification accuracy and their adaptation based on the class distributions of the data. It was shown that the SVSA method has competitive classification performance in comparison to the linear and nonlinear SVM with real world data.

During the implementation, it was observed that linear SVM gives the best classification performance if the data is linearly separable. In order to improve our algorithm, we plan to develop a hybrid algorithm that uses both linear SVM and

#### ACKNOWLEDGMENT

The authors would like to acknowledge the Scientific and Technological Research Council of Turkey (TUBITAK) for funding our research.

#### REFERENCES

- [1] G. A. Shmilovici, "The Data Mining and Knowledge Discovery Handbook", Springer, 2005.
- [2] Yue Shihong Li Ping Hao Peiyi, "Svm Classification :Its Contents and Challenges", Appl. Math. J. Chinese Univ. Ser. B, vol. 18 (3), 332-342, 2003.
- [3] C.C. Chang and C. Lin, "LIBSVM: A Library for Support Vector Machines", <http://www.csie.ntu.edu.tw/~cjlin/libsvm/>, 2001.
- [4] T. Cover and P. Hart, "Nearest neighbor pattern classification," IEEE Transactions on Information Theory, vol. 13 (1), pp.21-27, 1967.
- [5] T. Kohonen, "Learning vector quantization for pattern recognition," Tech. Rep., TKK-F-A601, Helsinki University of Technology, 1986.
- [6] T. Kohonen, J. Kangas, J. Laaksonen, and K. Torkkola, "Lvsqpk: A software package for the correct application of learning vector quantization algorithms," Neural Networks, IJCNN., International Joint Conference, vol. 1, pp. 725 - 730, 1992.
- [7] N. G. Kasapoğlu and O. K. Ersoy, "Border Vector Detection and Adaptation for Classification of Multispectral and Hyperspectral Remote Sensing", IEEE Transactions on Geoscience and Remote Sensing, Vol: 45-12, pp: 3880-3892, December 2007.
- [8] F. Melgani and L. Bruzzone, "Classification of hyperspectral remote sensing images with support vector machines", IEEE Transactions on Geoscience and Remote Sensing, vol. 42, no. 8, 2004.
- [9] R. Courant and D. Hilbert, "Methods of Mathematical Physics", Interscience Publishers, 1953.
- [10] J. A. Benediktsson, P. H. Swain, O. K. Ersoy, "Neural Network Approaches versus Statistical Methods in Classification of Multisource Remote Sensing-Data," IEEE Transactions Geoscience and Remote Sensing, Vol. 28, No. 4, pp. 540-552, July 1990.
- [11] S. Kaya, P. J. Curran and G. Llewellyn, "Post-earthquake building collapse: a comparison of government statistics as estimates derived from SPOT HRVIR data", International Journal of Remote Sensing, vol. 46, no. 13, pp. 2731-2740, 2005.