HYBRID SVM AND SVSA METHOD FOR CLASSIFICATION OF REMOTE SENSING IMAGES

G. Taşkın Kaya[1], *O. K. Ersoy*[2], *M. E. Kamaşak*[3]

[1]İstanbul Technical Univ, Informatics Institute, Istanbul, Turkey, e-mail: gulsen@be.itu.edu.tr
[2]Purdue University, School of Electrical and Computer Eng., Indiana, USA, e-mail: ersoy@purdue.edu
[3]İstanbul Technical Univ, Computer Engineering, Istanbul, Turkey, e-mail: kamasak@itu.edu.tr

ABSTRACT

A linear support vector machine (LSVM) is based on determining an optimum hyperplane that separates the data into two classes with the maximum margin. The LSVM typically has high classification accuracy for linearly separable data. However, for nonlinearly separable data, it usually has poor performance. For this type of data, the Support Vector Selection and Adaptation (SVSA) method was developed, but its classification accuracy is not very high for linearly separable data in comparison to LSVM. In this paper, we present a new classifier that combines the LSVM with the SVSA, to be called the Hybrid SVM and SVSA method (HSVSA), for classification of both linearly and nonlinearly separable data and remote sensing images as well. The experimental results show that the HSVSA has higher classification accuracy than the traditional LSVM, the nonlinear SVM (NSVM) with the radial basis kernel, and the previous SVSA.

***Index Terms*—** Support Vector Machines, Support Vector Selection and Adaptation, Hybrid SVM and SVSA.

1. INTRODUCTION

SVM is a binary classifier designed for two linearly separable classes by optimally finding a hyperplane with a maximum margin in the feature space [1]. SVM works well for linearly separable data but not necessarily for nonlinearly separable data.

Nonlinear SVM (NSVM) was also introduced in order to classify nonlinearly separable data with high accuracy. NSVM finds a nonlinear function by using a nonlinear kernel used to transform the current feature space into a high dimensional feature space. Although NSVM can achieve higher classification performance, it needs a high computation time for mapping data into the feature space [2]. In addition to consuming much time, it also requires optimized kernel parameters in order to obtain high classification accuracy. In many applications, the structure of the data is not known in advance.

For classifying such nonlinearly separable data, a novel method called Support Vector Selection and Adaptation (SVSA) was introduced in order to overcome NSVM's drawbacks with competitive performance [3]. The SVSA has less computation time compared to NSVM, and no kernels are needed.

The Hybrid SVSA and SVM (HSVSA) is presented here to make use of the best properties of both linear SVM and SVSA for high classification performance with all data. The hybrid model works as follows: in the nonlinear case, most data especially located near the hyperplane are missclassified with the Linear SVM. Since the SVSA is a nonlinear classifier, it is more effective to classify such data. Thefore, in the hybrid model, the SVSA method is used to classify such data. In the linear case, the data located near the hyperplane are classified with higher accuracy with the LSVM as compared to the SVSA. For such data, linear SVM is used in classification. In order to exploit the hybrid model, first the regions, where both SVM and SVSA has the highest classification performance need to be determined by using perpendicular distance to the hyperplane.

The paper consists of 4 sections. The SVSA and the HSVSA methods are presented in Sections 2 and 3, respectively. As a remote sensing implementation, earthquake damage assessment with the post earthquake Bam satellite image by the proposed method in comparison to other methods is presented in Section 4. Finally, conclusions are presented in Section 5.

2. SUPPORT VECTOR SELECTION AND ADAPTATION

The SVSA method consists of two stages: selection of support vectors obtained by LSVM and adaptation of the selected support vectors [4]. In the selection stage, some of the support vectors are eliminated as they are not sufficiently useful for classification. In the second stage, the remaining support vectors are adapted with respect to the training data to generate the reference vectors, which are subsequently used for classification of testing data. In terms of classification performance, the SVSA usually outperforms the LSVM. It is also competitive with NSVM in the classification of nonlinearly separable data. Therefore, a high classification accuracy can be achieved by the SVSA without the need for a kernel.

Let M, N, and J denote the number of training samples, the number of features, and the number of support vectors, respectively. Let $X = \{\mathbf{x}_1, \ldots, \mathbf{x}_M\}$ represent the training data with $\mathbf{x}_i \in \mathbf{R}^N$, Y represent the class labels with $y_i \in \{-1, +1\}$, and $S \in \{\mathbf{s}_1, \ldots, \mathbf{s}_J\}$ represent the support vectors with $\mathbf{s}_i \in \mathbf{R}^N$. Then, the linear SVM is employed to obtain the support vectors S from the training data X represented by

$$S = \left\{ \mathbf{s}_j \mid \mathbf{s}_j \in X \quad 1 \le j \le J \right\} \tag{1}$$

The training dataset T is updated to exclude the support vectors:

$$T = \left\{ \mathbf{t}_k \mid \mathbf{t}_k \in X \backslash S,\ k = 1, \ldots, N \right\} \tag{2}$$

In the selection stage, the support vectors in the set S are first classified by using the set T with the KNN algorithm [5]. The leave-one-out algorithm is used to determine the size of the neighborhood, K, for KNN classification. The result for $K = 1$ is given by

$$y^p_{s_j} = \left\{ y_{t_l} \mid l = \arg\min_k \left\{ \|\mathbf{s}_j - \mathbf{t}_k\| \right\},\ s_j \in S,\ t_k \in T \right\} \tag{3}$$

where $y^p_{s_j}$ is the predicted class label of the j^{th} support vector. If the original label and the predicted label of a support vector are different, then this support vector is excluded from the set of support vectors. According to the experiments conducted, the elimination of some support vectors by this process makes classification more accurate.

The remaining support vectors are called reference vectors and constitute the set R:

$$R = \left\{ \mathbf{r}_j \mid \mathbf{r}_j = \mathbf{s}_j \in S \quad \text{and} \quad y^p_{s_j} = y_{s_j} \right\} \tag{4}$$

The reference vectors are iteratively adapted based on the training data in a way to make them more representative for classification of data by the nearest neighbor rule [6]. The main logic of adaptation is that a reference vector causing a wrong decision should be further away from the current training vector. Similarly, the nearest reference vector with the correct decision should be closer to the current training vector.

Adaptation is achieved by using the Learning Vector Quantization (LVQ1) algorithm [7]. It is assumed that $\mathbf{r_j(t)}$ is the nearest reference vector to x_j with label y_{r_w}. The adaptation is applied as follows:

$$\mathbf{r_j}(t+1) = \begin{cases} \mathbf{r_j}(t) - \eta(t)(\mathbf{t_k} - \mathbf{r_j}(t)) & \text{if } y_{t_k} \neq y_{r_j} \\ \mathbf{r_j}(t) + \eta(t)(\mathbf{t_k} - \mathbf{r_j}(t)) & \text{if } y_{t_k} = y_{r_j} \end{cases} \tag{5}$$

It means that if the class label of the reference vector $\mathbf{r_j}$ (reference vector winner) matches the class label of the training sample $\mathbf{t_k}$, then the reference vector is moved towards $\mathbf{t_k}$. Otherwise, it is moved away from the input sample, where $0 \le \eta(t) \le 1$ is the corresponding learning rate parameter given by

$$\eta(t) = \eta_0 \left(1 - \frac{t}{T}\right) \tag{6}$$

where η_0 is the initial value of η, and T is the total number of learning iterations.

The adaptation is an iterative process and finds the adapted support vectors, called reference vectors, to be used for classification of the testing data by the nearest neighbor rule. In the classification of testing data, unseen instances are classified by using 1NN with the finalized reference vectors. When a new query instance is entered, 1NN assigns the label of the closest reference vector as its label.

The SVSA is generated as a binary classifier, but it is also generalized to classify multiclass data, for example, by using one-against one approach [8].

3. HYBRID SVSA

The LSVM gives the highest classification accuracy for linearly separable data. According to the results obtained with some experiments done with both SVSA and SVM, it was observed that the SVSA as well as the NSVM are not much efficient classifiers with linearly separable data compared to the linear SVM [9]. In order to increase the classification accuracy of the SVSA for linearly separable data and to generalize the SVSA method, the HSVSA method is introduced.

During the implementation of the SVSA, the results of the linear SVM are already available, and by utilizing this information, the hybrid model is generated by using the results of both LSVM and SVSA. Since the linear SVM is a binary classification method, the hybrid model is generated as a binary classifier and also generalized for the multiclass data, for example, by using the one against one approach.

In the HSVSA, first the training data is randomly partioned into k sets. A single set is retained as a validation dataset in order to determine the winner classifier between the SVM and the SVSA. The remaining $k - 1$ sets are used as training data to determine the separating hyperplane and the reference vectors. Afterwards, the perpendicular distances from each data in the validation set to this hyperplane are calculated and normalized. By using these distances, the feature space of the validation set are equally divided into n regions by using the hyperplanes generated parallel the LSVM hyperplane, as shown in Figure 1.

The classification accuracy of each method is calculated within each region, and the winner classifier having the highest classification accuracy is determined. In the classification of the testing data, the data lying within each region are determined by using their perpendicular distances to the separating hyperplane and classified by the region winner classifier. This process is then repeated k times (the folds), and each of the k sets are used exactly once as the validation data. The finalized

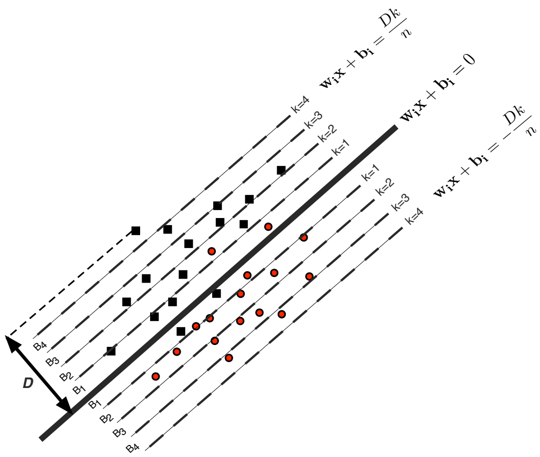

Fig. 1. HSVSA schema showing the partitioning of the decision space.

labels of the data are determined by using majority voting between k predicted labels. The algorithm of the HSVSA is shown in Algorithm 1.

Algorithm 1 HSVSA Algorithm

Inputs: $X = \{\mathbf{x_i} \mid \mathbf{x_i} \in \mathbf{R}^N\}_{i=1}^{M}$
Outputs $B_{i,t}$,region, and the winner classifier
Parameters: k *# validation sets*
Parameters: n *# regions*
for $i = 1$ to k **do**
  $V_i = X_i$ *Validation set*
  $\bar{X}_i = X \setminus X_i$ *Sub training set*
  Linear SVM
  $f_i(\mathbf{x}) = \mathbf{w}_i\mathbf{x} + \mathbf{b}_i, \quad \mathbf{x} \in \bar{X}_i$
  Support Vectors
  $S_i = \left\{\mathbf{s}_j, \mid \mathbf{s}_j \in \bar{X}_i\right\}, \ j \leq n$
  Run SVSA on the set, $\bar{X}_i$
  Reference vectors
  $R_i = \{\mathbf{r}_1, \ldots, \mathbf{r}_p\}, p \leq j$
  normalized distance to $f_i(\mathbf{x})$
  $d_i = \frac{|\mathbf{w_i}\mathbf{x}, +\mathbf{b_i}|}{\|\mathbf{w_i}\|^2}, \ d_i = \frac{d_i}{\|\mathbf{d}\|}, \quad \mathbf{x} \in V_i$
  Partitioning
  $B_{i,t} \equiv \left\{\mathbf{x} \mid \frac{d_i(t-1)}{n} \leq \frac{|\mathbf{w_i}\mathbf{x}+\mathbf{b_i}|}{\|\mathbf{w_i}\|^2} \leq \frac{d_i t}{n}\right\}_{t=1}^{n}, \mathbf{x} \in V_i$
  $\bar{\mathbf{y}}_{i,t}^{\text{SVM}} = \text{sign}\left(\mathbf{w}_i\mathbf{x} + \mathbf{b}_i\right), \quad \mathbf{x} \in B_{i,t}$
  $y_j^{\text{SVSA}} = y_t, t = \text{arg min}_l \parallel \mathbf{x}_j - \mathbf{r}_l \parallel, \mathbf{r}_l \in R_i$
  $\bar{\mathbf{y}}_{i,t}^{\text{SVSA}} = \left[y_1^{\text{SVSA}}, \ldots, y_q^{\text{SVSA}}\right]^T, \quad 1 \leq q \leq j$
  $\mathbf{y}_{i,t}$ *the labels for the region,* $B_{i,t}$
  classification errors for the region, $B_{i,t}$
  $E_{i,t}^{\text{SVM}} = \frac{1}{2} \parallel \mathbf{y}_{i,t} - \bar{\mathbf{y}}_{i,t}^{\text{SVM}} \parallel$
  $E_{i,t}^{\text{SVSA}} = \frac{1}{2} \parallel \mathbf{y}_{i,t} - \bar{\mathbf{y}}_{i,t}^{\text{SVSA}} \parallel$
  if $E_{i,t}^{\text{SVM}} < E_{i,t}^{\text{SVSA}}$ **then**
    SVM is the winner
  else
    SVSA is the winner
  end if
end for

4. EARTHQUAKE DAMAGE ASSESSMENT

As a remote sensing application, a post earthquake Quickbird satellite image was used to identify the damage patterns in the city of Bam, Iran during the 2003 earthquake. The HSVSA method, the LSVM, the SVSA and the NSVM were used for classification of the damaged and undamaged buildings. According to the results obtained, the hybrid model gave the best classification results in comparison to all the other methods.

The test area chosen within the city of Bam to detemine the earthquake damage is shown in Figure 2.

Fig. 2. Pre and post earthquakes image from area of interest, Iran Bam.

From the post earthquake image, vegetation, shadow, urban, open ground and damage area were chosen to be classified by the methods. The training and testing data was chosen by using MultiSpec software, and the numbers of data are tabulated in Table 1.

Class	Training	Testing
Vegetation	358	616
Shadow	276	449
Urban	356	866
Open Area	537	480
Damage	397	1237
Total	**1924**	**3648**

Table 1. Number of training and testing data.

Classification accuracies obtained with the methods are tabulated in Table 2.

According to Table 2, the overall classification accuracy of the SVSA method was increased from %82.3 to %83.8 with the HSVSA. The classsification accuracies of the linear SVM for the urban and open ground classes are higher

CLASSES	METHODS			
	SVM	SVSA	HSVSA	NSVM
Vegetation	99.5	**99.7**	**99.7**	99.5
Shadow	**99.4**	97.3	97.3	96.7
Urban	**91.8**	79.8	84.8	88.1
Open Ground	62.7	58.3	**60.9**	59.6
Damage	33.5	**79.3**	**79.3**	71.6
Overall Acc.	69.8	82.3	**83.8**	81.7

Table 2. Classification accuracies of the methods.

than those of the SVSA. With the HSVSA, their accuracies were also increased. Some statistical measures were also determined for especially measuring the damage pattern's separability from the other classes. Table 3 shows the kappa value, the true false positive and the false positive rates with each method for the damage class.

Methods	Statistical Results [%]		
	Kappa	TPR	FPR
SVM	62.3	67.1	29.6
SVSA	76.9	74.2	13.1
HSVSA	**78.9**	**77.7**	**12.9**
NSVM	76.3	77.3	16.3

Table 3. Statistical evaluation of testing results for each method.

According to Table 3, the HSVSA has the highest kappa value and true positive rate, and the lowest false positive rate as compared to the other methods. With the data set used, the HSVSA method was the best method to classify the damage pattern as compared to all the other methods.

The whole image was classified by the HSVSA and the resulting thematic map is shown in Figure 3.

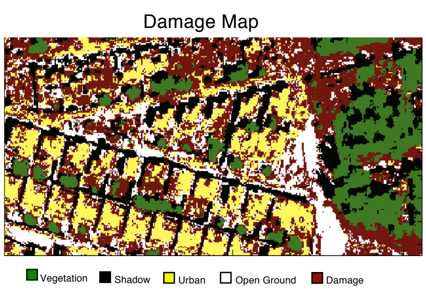

Fig. 3. The thematic map obtained with the HSVSA method.

5. CONCLUSIONS

The HSVSA method has been introduced to especially generalize the SVSA algorithm for classification of both linearly and nonlinearly separable data. The hybrid schema is based on partioning the decision space into quantized regions by using k fold validation. For each validation set, each region was labeled by the winner between the SVM and the SVSA method with respect to the highest classification performance. Partitioning was achieved by using the distances to the LSVM hyperplane. In the classification stage, the data falling in each region are classified by the region winning method. The final classification is achieved by using majority voting between k predicted values due to k fold validation. According the remote sensing data used, the HSVSA method had the best classification results as compared to LSVM, the SVSA, and, the NSVM.

6. REFERENCES

[1] V. Cherkassky and F. Mulier, *Learning From Data: Concepts, Theory, and Methods*, Wiley-interscience, New York, 1998.

[2] S. Yue, P. Li and P. Hao, "Svm classification:its contents and challenges," *Appl. Math. Chin. Univ.*, vol. 18(3), pp. 332–342, Sep. 2003.

[3] G. Taskin Kaya, O. K. Ersoy and M. E. Kamasak, "Support vector selection and adaptation for classification of earthquake images," in *IEEE, International Geoscience and Remote Sensing Symposium*, July 12-17, 2009.

[4] G. Kaya, O. K. Ersoy and M. E. Kamasak, "Support vector selection and adaptation and its application in remote sensing," in *4th International on Recent Advances in Space Technologies*. Jun. 2009, pp. 408–412, IEEE.

[5] T. Cover and P. Hart, "Nearest neighbor pattern classification," *IEEE Trans. Inf. Theory*, vol. IT-13, pp. 21–27, Jan. 1967.

[6] N. G. Kasapoglu, O. K. Ersoy , "Border vector detection and adaptation for classification of multispectral and hyperspectral remote sensing images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 45, no. 12, pp. 3880–3892, Dec. 2007.

[7] T. Kohonen, "Learning vector quantization for pattern recognition," Tech. Rep., TKK-F-A601, Helsinki University of Technology, 1986.

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

[9] G. Taskin Kaya, O. K. Ersoy and M. E. Kamasak, "Support vector selection and adaptation," in *International Symposium on Innovations in Intelligent Systems and Applications*, June 29-July 1, 2009.