

SUPPORT VECTOR SELECTION AND ADAPTATION FOR CLASSIFICATION OF EARTHQUAKE IMAGES

G. Taşkın Kaya^{*1}, O. K. Ersoy², M. E. Kamaşak³

¹Istanbul Technical Univ, Informatics Institute, Istanbul, Turkey, e-mail: gulsen@be.itu.edu.tr

²Purdue University, School of Electrical and Computer Eng., Indiana, USA, e-mail: ersoy@purdue.edu

³Istanbul Technical Univ, Computer Engineering, Istanbul, Turkey, e-mail: kamasak@itu.edu.tr

ABSTRACT

In this paper, we propose a new machine learning algorithm that we named Support Vector Selection and Adaptation (SVSA). Our aim is to achieve the classification performance of the nonlinear support vector machines (SVM) by using only the support vectors of the linear SVM. The proposed method does not require any type of kernels, and requires less computation time compared to the nonlinear SVM. The SVSA algorithm has two steps: selection and adaptation. In the first step, some of the support vectors obtained from linear SVM are selected. Then the selected support vectors are adapted iteratively in the training algorithm. The proposed method are compared against the linear and nonlinear SVM on synthetic and real remote sensing data. The results show that the proposed SVSA algorithm achieves very close performance to nonlinear SVM without any kernels in less computation time.

Index Terms— Support Vector Machines, Support Vector Selection and Adaptation, Classification of Earthquake Images

1. INTRODUCTION

Linear support vector machine (SVM) is based on determining an optimum hyperplane that separates the data into two classes with the maximum margin [1, 2]. Linear SVM typically have high classification accuracy for linearly separable data. However, for nonlinearly separable data linear SVM has poor performance. For this type of data, nonlinear SVMs are preferred. Nonlinear SVMs transform input data using a nonlinear kernel followed by regular SVM.

Although nonlinear SVM can achieve higher classification performance, it requires high computation time to map the input to a higher dimensional space by a nonlinear kernel function which is a fully dense matrix [3]. Computational complexity of nonlinear SVM grows with the cube of the total number of training data, $O(n^3)$, whereas it is $O(n^2)$ for

linear SVM. Furthermore, the selection of this nonlinear kernel requires some a priori information about the structure of the data. In many applications, the structure of the data is not known. Therefore, kernel selection might be a challenging task. After the selection of the kernel for nonlinear SVM, the kernel parameters have to be adjusted for maximum performance. The optimal kernel parameters can be found by cross validation procedures. In summary, the kernel selection and kernel parameters are critical for the nonlinear SVM performance.

Support Vector Selection and Adaptation (SVSA) was introduced to overcome the mentioned drawbacks of nonlinear SVM without a significant performance loss [4]. SVSA method has some advantages over linear and nonlinear SVM. It requires less computation time compared to nonlinear SVM and no kernels are needed. On the nonlinearly separable data, classification performance of SVSA is very close to nonlinear SVM.

2. SUPPORT VECTOR SELECTION AND ADAPTATION

The SVSA method consists of two stages: Selection of support vectors from the training data and adaptation of the selected support vectors. In the selection stage, some of the support vectors are eliminated as they are not sufficiently useful for classification. After the elimination, the remaining support vectors are adapted and used as reference vectors for classification. In this way, nonlinear classification is achieved without need for a kernel.

Let M , N , and J denote the number of training samples, the number of features, and number of support vectors respectively. Let $X = \{\mathbf{x}_1, \dots, \mathbf{x}_M\}$ represent the training data with $\mathbf{x}_i \in \mathbf{R}^N$, $Y \in \mathbf{R}^M$ represent the class labels with $y_i \in \{-1, +1\}$, and $S \in \{\mathbf{s}_1, \dots, \mathbf{s}_J\}$ 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) as follows:

$$S = \{(\mathbf{s}_j, y_{s_j}) \mid (\mathbf{s}_j, y_{s_j}) \in X \quad 1 \leq j \leq J\} \quad (1)$$

^{*}Thanks to The Scientific and Technological Research Council of Turkey (TUBITAK) for funding.

where $y_{s_j} \in \{-1, +1\}$ is the class label of the j^{th} support vector.

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

$$T = \{(\mathbf{t}_k, y_{t_k}) \mid (\mathbf{t}_k, y_{t_k}) \in X \setminus S, k = 1, \dots, N - J\} \quad (2)$$

In the selection stage, the labels of the support vectors in the set S are reassigned with respect to the set, T , by using the K-Nearest Neighbor (KNN) algorithm [5].

$$y_{s_j}^p = \{y_{t_l} \mid l = \arg \min_k \{\|\mathbf{s}_j - \mathbf{t}_k\|\}, s_j \in S, t_k \in T\}$$

where $y_{s_j}^p$ 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 eliminated. The remaining support vectors are called reference vectors and constitute the set R :

$$R = \{(\mathbf{r}_j, y_{r_j}) \mid (\mathbf{s}_j, y_{s_j}) \in S \text{ and } y_{s_j}^p = y_{s_j}\} \quad (3)$$

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

The reference vectors are iteratively adapted based on the training data in a way to increase the distance between neighboring reference vectors with different class labels [6]. The main idea of the adaptation is that a reference vector causing a wrong decision should be further away from the current training vector, and the nearest reference vector with the correct class should be closer to the current training vector. Adaptation is achieved by using a method that has some similarities to the Learning Vector Quantization (LVQ) algorithm [7].

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

$$\mathbf{r}_w(\mathbf{t} + 1) = \begin{cases} \mathbf{r}_w(\mathbf{t}) - \eta(\mathbf{t})(\mathbf{x}_j - \mathbf{r}_w(\mathbf{t})) & \text{if } y_j \neq y_{r_w} \\ \mathbf{r}_w(\mathbf{t}) + \eta(\mathbf{t})(\mathbf{x}_j - \mathbf{r}_w(\mathbf{t})) & \text{if } y_j = y_{r_w} \end{cases} \quad (4)$$

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

$$\eta(t) = \eta_0 e^{-t/\tau} \quad (5)$$

where η_0 is the initial value of η and τ is a time constant.

The adaptation is an iterative process and finds the reference vectors to be used for classification of the data. The iteration is terminated when the learning rate is less than a predetermined value.

The adapted reference vectors are used for classification of the training and testing sets. For this purpose, the 1-NN method is applied to classify the samples with respect to these reference vectors. The Euclidian distances from the input vector to the reference vectors are calculated, and classification is done based on closest reference vector's label.

3. EXPERIMENTS WITH SYNTHETIC AND EARTHQUAKE DATA

In the first experiment, we first generated different types of synthetic data with different types of nonlinearity in order to compare the classification performance of the proposed method with the nonlinear SVM [8].

In the second experiment, post-earthquake Quickbird satellite image with high resolution was used to identify damage patterns in a small area of Bam, Iran during the 2003 earthquake.

3.1. Experiment 1: Synthetic Data

The synthetic data with four different distributions, each of which has two features and two classes, were produced (Fig. 1). The synthetic data helps us to analyze the classification performances of all algorithms on data with different distributions. Ten subsets of training and testing data for each dataset are randomly chosen by 40 % and 60 % from whole the data, respectively.

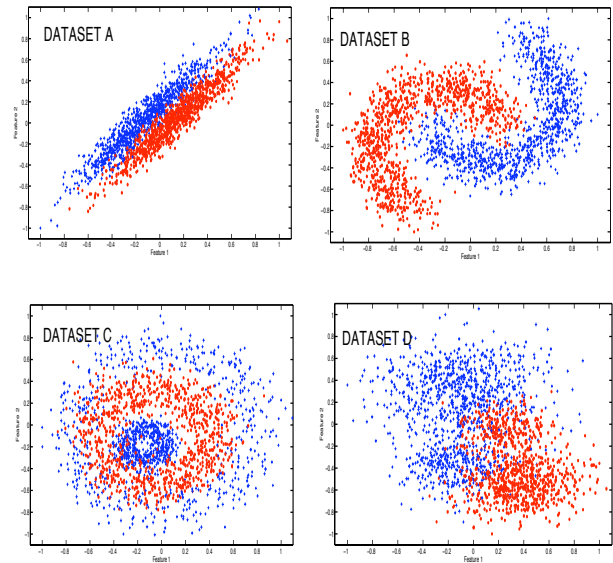


Fig. 1. Distributions of Synthetic Data.

All algorithms are used to classify these ten subsets of data for each data distribution. The average classification accuracy of each method is given in Table 1. In Table 1, A, B, C, and D are the synthetic data with different distributions that are shown in Fig. 1, and NSVM-1 and NSVM-2 refer to nonlinear SVM with radial basis kernel function and polynomial kernel function, respectively.

It can be seen from Table 1 that the classification performance of the SVSA is better than the linear SVM except in data type A. In addition, the SVSA has quite close classification performance to nonlinear SVM with radial basis kernel

Table 1. The mean and standard deviation of the classification performance for the synthetic data. NSVM-1 and NSVM-2 refers to nonlinear SVM with radial basis kernel function and polynomial kernel function, respectively.

Mean				
Method	Types of Data			
	A	B	C	D
SVM	92.69	85.08	50.37	85.82
SVSA	91.83	97.35	85.43	88.32
NSVM-1	92.43	98.13	86.51	88.85
NSVM-2	92.65	80.45	50.12	85.59
Standard Deviation				
Method	Types of Data.			
	A	B	C	D
SVM	0.55	0.82	4.96	0.66
SVSA	0.64	0.45	1.15	0.89
NSVM-1	0.46	0.35	1.71	0.95
NSVM-2	0.39	0.96	4.96	0.65

function, and it has better classification performance than the nonlinear SVM with polynomial kernel function. It is also worthwhile to note that the classification by nonlinear SVM done with different type of kernels yields the different classification performances. Moreover, the stability of the SVSA is quite close to NSVM-1.

From the results obtained, it can be inferred that as the data have more nonlinear distribution, the SVSA performance gets closer to nonlinear SVM.

3.2. Experiment 2: Earthquake Data

The post earthquake image is used in select the samples to train and test the algorithms. These samples are classified into four classes: Damage, buildings, vegetation and open ground (Fig. 3).

One hundred samples are randomly chosen by 40 % and 60 % for training and testing. All the methods were used to classify four classes over the one hundred samples, and the overall classification performances of each method were compared.

The average classification performance of the algorithms and their standard deviations are shown in Fig. 2. From this figure, SVSA gives the highest classification performance compared to both linear and nonlinear SVM. The standard deviation of the SVSA algorithm is also the smallest amongst all classification algorithms.

SVSA is used for classification of whole area of interest over the post earthquake image (Fig. 2).

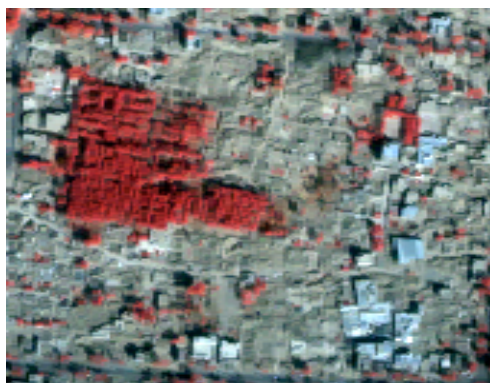


Fig. 3. Post earthquake image for area of interest.

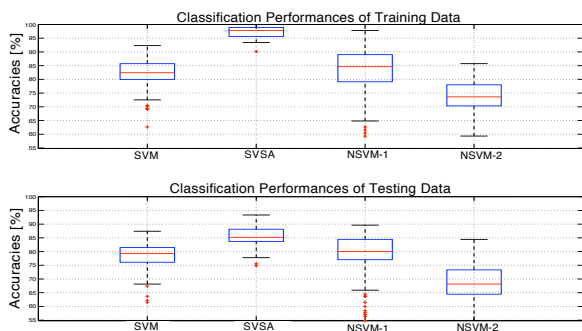


Fig. 2. The classification accuracies of all the methods over the one hundred samples.

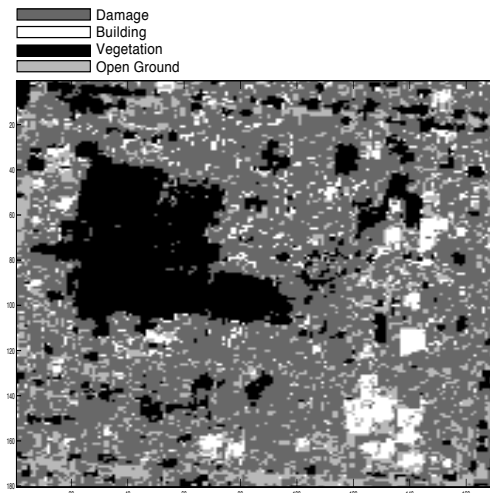


Fig. 4. The thematic map of the classes obtained by the SVSA algorithm.

All the classes were detected by using the SVSA method as in Figure 4. The SVSA took less time in comparison to the nonlinear SVM during the classification.

4. CONCLUSION

In this paper, we proposed a novel support vector selection and adaptation method which are reliable for both linearly separable and nonlinearly separable data. The SVSA method consists of selection of support vectors which most contribute to the classification accuracy and adaptation of them based on the class distributions of the data. The proposed algorithm is tested on synthetic and remote sensing data. The classification performance is compared against linear and nonlinear SVM algorithms. It is shown that the SVSA method gives better classification results compared to linear SVM on nonlinearly separable data, and it gives satisfactory classification performance in comparison to the nonlinear SVM for both synthetic and real data.

5. REFERENCES

- [1] V. Cherkassky and F. Mulier, *Learning From Data : Concepts, Theory and Methods*, Wiley-interscience, 1998.
- [2] C.C. Chang and C. Lin, *LIBSVM: a library for support vector machines*, 2001.
- [3] Yue Shihong, Li Ping, Hao Peiyi, "Svm classification:its contents and challenges," *Appl. Math. J. Chinese Univ. Ser. B*, vol. 18(3), pp. 332–342, 2003.
- [4] G. Taşkın Kaya, O. K. Ersoy, "Support vector selection and adaptation for classification of remote sensing images," *Purdue University Technical Report, TR-ECE-09-2*, 2008.
- [5] T. Cover and P. Hart, "Nearest neighbor pattern classification," *IEEE Transactions on Information Theory*, vol. 13(1), pp. 21–27, 1967.
- [6] N. G. Kasapoğlu, 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, 2007.
- [7] T. Kohonen, "Learning vector quantization for pattern recognition," Tech. Rep., TKK-F-A601, Helsinki University of Technology, 1986.
- [8] R. P. W. Duin, P. Juszczak, P. Paclik and et. al, "A matlab toolbox for pattern recognition," *Delft University of Technology, PRTools4.1*, 2007.