

Review: Support Vector Machines in Pattern Recognition

Parashjyoti Borah ^{#1}, Deepak Gupta ^{*2}

[#]Electronics & Computer Engineering, NIT, Arunachal Pradesh, Yupia, India
¹parashjyoti@hotmail.com

^{*}Electronics & Computer Engineering, NIT, Arunachal Pradesh, Yupia, India
²deepak.cse@nit.ap.in

Abstract— SVM is extensively used in pattern recognition because of its capability to classify future unseen data and its' good generalization performance. Several algorithms and models have been proposed for pattern recognition that uses SVM for classification. These models proved the efficiency of SVM in pattern recognition. Researchers have compared their results for SVM with other traditional empirical risk minimization techniques, such as Artificial Neural Network, Decision tree, etc. Comparison results show that SVM is superior to these techniques. Also, different variants of SVM are developed for enhancing the performance. In this paper, SVM is briefed and some of the pattern recognition applications of SVM are surveyed and briefly summarized.

Keyword - Hyperplane, Pattern Recognition, Quadratic Programming Problem, Support Vector Machines.

I. INTRODUCTION

Support Vector Machine (SVM) is a classification technique developed by Cortes & Vapnik [4], originally designed for binary classification. While traditional methods of pattern recognition try to minimize the empirical risk, SVM tries to minimize structural risk [4]. SVM tries to minimize the upper bound of generalization error by maximizing the distance between the separating hyperplanes and the data points. It condenses the information from the training data that makes classification much faster than the traditional empirical risk minimization techniques. Because of its generalization capability even in higher dimension, SVM is used in many pattern recognition areas, such as object detection [6],[19-22], handwritten character recognition [5],[7],[23], face recognition [8],[12],[15] and speech recognition [2],[9],[10],[13],[24]. In this paper, we have briefly summarized about SVM and presented a brief overview of some of the application of SVM in pattern recognition.

II. SUPPORT VECTOR MACHINE

SVM estimates a $f : R^n \rightarrow \{\pm 1\}$ function with the help of training patterns x_i and class labels $y_i \in \{-1,+1\}$ where the training patterns are n-dimensional, i.e. $(x_1, y_1), (x_2, y_2), \dots, (x_m, y_m) \in R^n \times \{\pm 1\}$. Each of the data points x_i belong to either of the two classes labelled $y_i \in \{+1,-1\}$. The goal is to define a hyperplane that maximizes the distance between the two class boundaries and divides all the input data into two classes, with all points of one class falling on one side of the hyperplane and of the other class falling on the other side.

A. Linearly Separable Case

In this case, for all the input data points, there exists at least one hyperplane, that linearly separates the two classes and the data points of each class fall into respective class in space. The goal is to find out the hyperplane that maximizes the distance between the two class boundaries. The data points of each class closest to a hyperplane are called Support Vectors. Fig .1 depicts Linear SVM for separable case. Let us define,

$$w.x_i + b \geq 1, \text{ if } y_i = +1$$

$$w.x_i + b \leq -1, \text{ if } y_i = -1$$

for $i = 1, 2, \dots, m$.

The above set of equations can be generalized as,

$$y_i(w.x_i + b) \geq 1 \tag{1}$$

The equation for a separating hyperplane (w, b) can be derived as

$$w.x + b = 0$$

The distance between the two marginal hyperplanes is

$$d = \frac{2}{\|w\|} \tag{2}$$

Hence, the optimum separating hyperplane can be considered as the solution to the problem of maximizing $1/\|w\|$ subject to constraint (1), which, for mathematical convention, can be formalized as,

$$\min \frac{1}{2} w' w$$

$$\text{subject to, } y_i(w.x_i + b) \geq 1, i = 1, 2, \dots, m \quad (3)$$

Instead of solving equation (3) in primal form, one can solve its dual form by using Lagrange's Multiplier [16] and apply the Karush-Kuhn-Tucker (KKT) conditions.

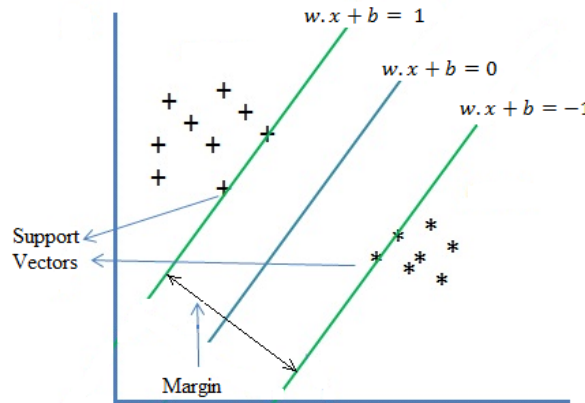


Fig. 1. SVM for linearly separable case

B. Linearly Non-Separable Case

But, in real-life data, the data points are not linearly separable in most of the cases. Though, a maximal margin hyperplane that minimizes misclassification error can be determined by introducing a positive slack variable ξ_i in constraint (3).

$$y_i(w.x_i + b) \geq 1 - \xi_i, i = 1, 2, \dots, m$$

The variable ξ_i exceeds unity in case of an error. $\sum_i \xi_i$ is the upper bound for total of misclassification error. Hence, the objective function in (3) becomes,

$$\min \frac{1}{2} w' w + C \sum \xi_i$$

$$\text{subject to, } y_i(w.x_i + b) \geq 1 - \xi_i, i = 1, 2, \dots, m \quad (4)$$

Here, C is a parameter chosen by the user. Solving the above problem of equation (4) gives the Generalized Optimum Separating Hyperplane. Fig.2 illustrates SVM for linearly non-separable case.

C. Non Linear Kernels

In most of the practical cases, linear separation of input data points gets too restrictive. The input space in those cases is mapped to a higher dimensional feature space using a kernel function where the feature vectors can be linearly separated by a hyperplane.

In order to accomplish separability, the input data are non-linearly mapped into a higher dimensional feature space such as $\Phi: R^n \rightarrow R^p$, where $P > n$. Fig. 3 depicts the mapping of input data sample into feature space by using Φ transformation. The training is then performed on the data obtained from the dot product $\Phi(x_i) \cdot \Phi(x_j)$. Although, Φ is not known a priori and the dot product of the mapping functions $\Phi(x_i) \cdot \Phi(x_j)$ is very expensive and complex. Using the Mercer's theorem [1] for positive definite functions $\Phi(x_i) \cdot \Phi(x_j)$ can be replaced by a symmetric kernel function $k(x_i, x_j)$ as follows,

$$k(x_i, x_j) = (\Phi(x_i) \cdot \Phi(x_j))$$

As it can be seen that it is not necessary to know the mapping function Φ in order to calculate the feature vectors. Knowing only the input data and the kernel function is enough to calculate the training data. Some popular kernel functions are stated below.[3]

1. Polynomial kernel, $k(x, x_i) = (x^T x_i + 1)^d$
2. Gaussian kernel, $k(x, x_i) = \exp(-\|x - x_i\|^2 / 2\sigma^2)$ where $\sigma > 0$ is a parameter.
3. Multi-layer perceptron (sigmoid), $k(x, x_i) = \tanh(\beta_0 x^T x_i + \beta_1)$, where β_0 and β_1 are user defined values and $i = 1, 2, \dots, m$.

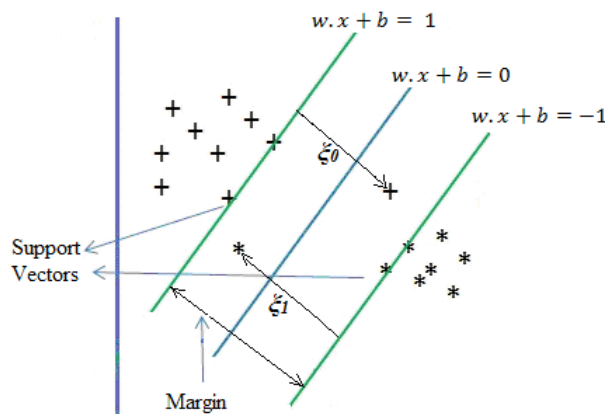


Fig. 2. SVM for linearly non-separable case

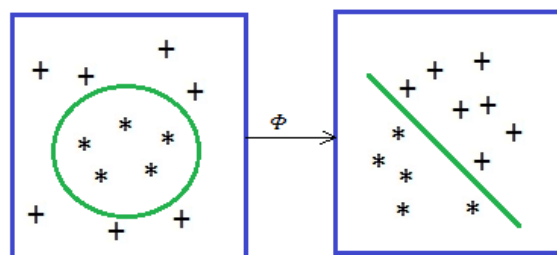


Fig. 3. Mapping of input space to feature space in linearly non-separable case

III. APPLICATIONS OF SUPPORT VECTOR MACHINE

SVMs are extensively used for pattern recognition. Researchers have proposed and developed many methods and techniques to solve pattern recognition problems using SVM. In this section, some existing methods of pattern classification are roughly categorized based on their purpose.

A. Object Detection And Recognition

Object detection is the technique that deals with detection of objects of certain class (eg. animals, trees, vehicles, etc.) in an image or a video. SVM comes handy in automatic detection of object of such classes when trained with proper training data. One such method was proposed and developed by Nakajima et al. [19]. They have defined a multi-class classification problem for people recognition and pose estimation. Authors have used pair-wise and DDAG multi-class SVMs with linear kernel. They have experimented on 640 images, taken 40 images for each of four different people selected. Each person was taken images with four different poses. For people and pose recognition, two features (colour histogram and local shape) were selected and are tested with. Results showed that the local shape feature outperformed the colour histogram feature.

Roobaert and Van Hulle [22] developed a model for recognizing 3D objects using SVMs. Their experiment was based on the COIL object database that contains 7200 images of 100 objects and each object with 72 different views. They have summarized their results with different number of training views taken and they found that with less than 18 training views taken, the performance of the model decreases. Pontil and Verri's [20] work is similar to [22]. They also have used COIL object database for their experiment. They have used linear SVMs for aspect-based 3D object recognition from a single view. Unlike [22], their experiment was performed without feature extraction, data reduction and pose estimation. Hence, the testing images contained noise, occlusion and pixel shifts. However, their result has shown a very good performance.

Pittore et al [21] proposed a system to detect the presence of moving people. They have represented the event by using a SVM for regression, and recognized trajectory of visual dynamic events from an image sequence by SVM classifier. Authors named their developed system as VIDERE (VISual Dynamic Event REcognition). Gao et al. [6] proposed an SVM based algorithm that tries to detect moving vehicles from shadows using shadow and head-lights elimination technique. The considered the problem to a simple two-class classification scenario.

B. Handwritten Character Recognition

In handwritten character recognition a computer receives input data from sources such as paper documents, photographs, touch-screens and other devices and interprets them as characters or letters or words of a language. SVM found to be very effective as compared to other learning algorithms in recognizing handwritten characters. A major problem in handwritten character recognition is its huge variability and distortions of pattern. Choisy and Belaid [5] proposed model to recognize French bank cheque words. For local view NSPH-HMM and for global view SVM were used by the authors.

Gorgevik et al. [7] used SVMs for handwritten digits recognition. They tested their model with single SVM classifier, and with two different SVM classifiers whose results were combined together using rule-based reasoning. Their experimental results shows, single SVM classifier is more efficient over rule-based reasoning in recognizing handwritten digits. Teow et al. [23] have proposed a digit recognition system that uses a linear SVM classifier by extracting features that are biologically plausible, linearly separable and semantically clear.

C. Face Recognition

Face recognition is a well-established field of research that deals with identifying or verifying a person from a digital image or a video frame from a video source. Many face recognition systems using SVM with very high performance, were developed till date. Among them, Guo et al. [8] proposed a multi-class SVM classifier for face recognition. They have also compared the results of their model with Nearest Center (NC), Hidden Markov Model (HMM), Conventional Neural Network (CNN), and Nearest Feature Line (NFL). The input dataset was first normalized using PCA. Then the normalized data were input to the SVM classifier. The SVM classifier outperformed the others with an error rate of 3% on ORL face database. In another model, Kim et al. [15] explored spatial relationship among potential eye, nose and mouth objects for face recognition using a modified SVM local correlation kernel. They compared their proposed kernel with existing kernels. It showed better performance as compared to the others with an error rate of 2% when tested on ORL database.

A component-based method was proposed by Heisele et al. [12] and its performance was compared with two global methods for face recognition by one-to-others SVMs. Huang et al. [13] proposed a model that generated a large number of synthetic face images for training of the system. The 3D models were rendered to train the system under various poses and illumination. In their component-based system, a single feature vector was formed by extracting and combining the facial components, which was then classified by the SVMs. The component-based method was compared to two global methods, which showed that component-based method performed better than the global methods.

D. Speaker Recognition And Speech Recognition

Discriminative classifiers and generative model classifiers are the two most popular techniques in speaker and speech recognition. SVMs are generally used in discriminative classifiers. SVMs were used by Bengio and Mariethoz [2] for speaker verification. They performed their experiments on different datasets. Instead of the classical thresholding rules, SVMs decide whether to accept or reject. Wan and Campbell [24] proposed a new technique in which they have normalized the traditional polynomial kernel and used with SVMs for speech recognition.

Some researchers applied SVM to visual speech recognition [9], [10]. Visual speech recognition recognizes speech from lip-reading of the speaker. In visual speech recognition, a particular sound uttered by the speaker is described by a generic facial image, called a viseme. Each viseme is described by SVM. SVMs were used as nodes by vitterbi algorithm for modelling the temporal character speech. Performance was evaluated by experimenting on audio-visual data Tulip 1, to solve the task of recognizing the first four digits in English [9], [10].

E. Some Other Applications

There are many other applications of SVM in pattern recognition problems. Nonlinear SVM was used by Moghaddam and Yang [17], [25] for gender classification. They have used FERET face dataset for their experiment. The training set of FERET dataset contains 1496 images (793 males and 713 females) and the test set contains 259 images (133 males and 126 females). They have used five –fold cross-validation technique for training and testing of each classifier using the face images. SVM outperformed some existing traditional classifiers with an error rate of 3.4%.

Gutta et al. [11] used SVM to classify face poses. FERET database was used for their experiment. Experimental results showed 100% accuracy. Huang et al. [14] also performed face pose detection using SVM. They classified face poses into three categories. Yao et al. [26] used multi-class SVM for fingerprint classification. Combining flat and structured representation of the features of fingerprints was used to train the SVMs. Results showed good performance of the model. Also, SVMs are widely used in detecting intrusions to a network or a host computer. In a model, proposed by S. Mukkamala [18], SVM light was used for designing intrusion detection system. Also, in comparison of SVM to ANN, SVM is found to be more efficient than ANN

in terms of training time and detection time. Experiment was performed on benchmark DARPA dataset and authors have claimed above 99% accuracy for both the systems.

IV. CONCLUSION

In this paper, we have presented briefly about SVM and discussed some applications of SVM in pattern recognition problems. Because of its excellent generalization performance SVMs are extensively used to solve various pattern recognition problems. Some of the applications are presented in this paper. Initially, SVM was applied for two-class classification problems. But, to obtain more specificity in results, SVMs are enhanced and used for multi-class classification problems as well. In some cases, different variants of the original SVM are also applied to yield better performance.

To prove superiority of SVM in pattern recognition problem, some authors have compared performance of SVM with other traditional empirical risk minimization techniques. In those cases, SVM outperformed those techniques in terms of training and testing time efficiency and classification accuracy due to having structural risk minimization principle. Some researchers have experimented SVM with different kernels and compared their performance. With proper selection of the kernel and balanced training data along with relevant feature selection, SVMs yield excellent result. SVMs are widely used in other classification problems also and proven to deliver excellent performance. It is an ongoing active field of research and lots of dimensions are yet to be unfolded. Although SVM performs excellent for classification problems but still it can be improved by some better techniques to find out the optimum parameters. Other future work can be proposed online learning based algorithm for real-life applications.

REFERENCES

- [1] Burges, Christopher JC. "A tutorial on support vector machines for pattern recognition." *Data mining and knowledge discovery* 2.2 (1998): 121-167.
- [2] Bengio, Samy, and Johnny Mariétoz. "Learning the decision function for speaker verification." *Acoustics, Speech, and Signal Processing, 2001. Proceedings.(ICASSP'01), 2001 IEEE International Conference on*. Vol. 1. IEEE, 2001.
- [3] Byun, Hyeran, and Seong-Whan Lee. "A survey on pattern recognition applications of support vector machines." *International Journal of Pattern Recognition and Artificial Intelligence* 17.03 (2003): 459-486.
- [4] Cortes, Corinna, and Vladimir Vapnik. "Support-vector networks." *Machine learning* 20.3 (1995): 273-297.
- [5] Choisy, Christophe, and Abdel Belaid. "Handwriting recognition using local methods for normalization and global methods for recognition." *Document Analysis and Recognition, 2001. Proceedings. Sixth International Conference on*. IEEE, 2001.
- [6] Gao, Dashan, Jie Zhou, and Leping Xin. "SVM-based detection of moving vehicles for automatic traffic monitoring." *Intelligent Transportation Systems, 2001. Proceedings. 2001 IEEE*. IEEE, 2001.
- [7] Gorgevik, Dejan, Dusan Cakmakov, and Vladimir Radevski. "Handwritten digit recognition by combining support vector machines using rule-based reasoning." *Information Technology Interfaces, 2001. ITI 2001. Proceedings of the 23rd International Conference on*. IEEE, 2001.
- [8] Guo, Guodong, Stan Z. Li, and Kapluk Chan. "Face recognition by support vector machines." *Automatic Face and Gesture Recognition, 2000. Proceedings. Fourth IEEE International Conference on*. IEEE, 2000.
- [9] Gordan, Mihaela, Constantine Kotropoulos, and Ioannis Pitas. "Application of support vector machines classifiers to visual speech recognition." *Image Processing, 2002. Proceedings. 2002 International Conference on*. Vol. 3. IEEE, 2002.
- [10] Gordan, Mihaela, Constantine Kotropoulos, and Ioannis Pitas. "Visual speech recognition using support vector machines." *Digital Signal Processing, 2002. DSP 2002. 2002 14th International Conference on*. Vol. 2. IEEE, 2002.
- [11] Srinivas Gutta, Jeffrey R. J. Huang, P. Jonathon and Harry Wechsler. "Mixture of experts for classification of gender, ethnic origin, and pose of human faces." *IEEE Transactions on neural networks* 11.4 (2000): 948-960.
- [12] Heisele, Bernd, Purdy Ho, and Tomaso Poggio. "Face recognition with support vector machines: Global versus component-based approach." *Computer Vision, 2001. ICCV 2001. Proceedings. Eighth IEEE International Conference on*. Vol. 2. IEEE, 2001.
- [13] Huang, Jennifer, Volker Blanz, and Bernd Heisele. "Face recognition using component-based SVM classification and morphable models." *Pattern Recognition with Support Vector Machines*. Springer Berlin Heidelberg, 2002. 334-341.
- [14] Huang, Jeffrey, Xuhui Shao, and Harry Wechsler. "Face pose discrimination using support vector machines (SVM)." *Pattern Recognition, 1998. Proceedings. Fourteenth International Conference on*. Vol. 1. IEEE, 1998.
- [15] Hyun-Chul Kim, Shaoning Pang, Hong-Mo Je, Daijin Kim and Sung Yang Bang. "Pattern classification using support vector machine ensemble." *Pattern Recognition, 2002. Proceedings. 16th International Conference on*. Vol. 2. IEEE, 2002.
- [16] Mangasarian, Olvi L., and David R. Musicant. "Lagrangian support vector machines." *Journal of Machine Learning Research* 1.Mar (2001): 161-177.
- [17] Moghaddam, Baback, and Ming-Hsuan Yang. "Learning gender with support faces." *IEEE Transactions on Pattern Analysis and Machine Intelligence* 24.5 (2002): 707-711.
- [18] Mukkamala, Srinivas, Guadalupe Janoski, and Andrew Sung. "Intrusion detection using neural networks and support vector machines." *Neural Networks, 2002. IJCNN'02. Proceedings of the 2002 International Joint Conference on*. Vol. 2. IEEE, 2002.
- [19] Nakajima, Chikahito, Massimiliano Pontil, and Tomaso Poggio. "People recognition and pose estimation in image sequences." *Neural Networks, 2000. IJCNN 2000, Proceedings of the IEEE-INNS-ENNS International Joint Conference on*. Vol. 4. IEEE, 2000.
- [20] Pontil, Massimiliano, and Alessandro Verri. "Support vector machines for 3D object recognition." *IEEE transactions on pattern analysis and machine intelligence* 20.6 (1998): 637-646.
- [21] Pittore, Massimiliano, Curzio Basso, and Alessandro Verri. "Representing and recognizing visual dynamic events with support vector machines." *Image Analysis and Processing, 1999. Proceedings. International Conference on*. IEEE, 1999.
- [22] Roobaert, Danny, and Marc M. Van Hulle. "View-based 3d object recognition with support vector machines." *Neural Networks for Signal Processing IX, 1999. Proceedings of the 1999 IEEE Signal Processing Society Workshop.. IEEE, 1999*.
- [23] Teow, Loo-Nin, and Kia-Fock Loe. "Robust vision-based features and classification schemes for off-line handwritten digit recognition." *Pattern Recognition* 35.11 (2002): 2355-2364.
- [24] Wan, Vincent, and William M. Campbell. "Support vector machines for speaker verification and identification." *Neural Networks Signal Process Proc IEEE*. Vol. 2. 2000.

- [25] Yang, Ming-Hsuan, and Baback Moghaddam. "Gender classification using support vector machines." *Image Processing, 2000. Proceedings. 2000 International Conference on*. Vol. 2. IEEE, 2000.
- [26] Yuan Yuan, Gian Luca Marcialis, Massimiliano Pontil, Paolo Frasconi and Fabio Roli. "Combining flat and structured representations for fingerprint classification with recursive neural networks and support vector machines." *Pattern Recognition* 36.2 (2003): 397-406.

AUTHOR PROFILE

Parashjyoti Borah is currently pursuing Ph.D. in Electronics and Computer Engineering from National Institute of Technology, Arunachal Pradesh, India. He received his Master of Technology degree in Computer Science and Engineering from Assam Don Bosco University, Azara, India in 2015. His main research includes support vector machines, fuzzy systems and other data mining techniques.

Deepak Gupta is an Assistant Professor of National Institute of Technology, Arunachal Pradesh, India. He received his Ph.D. from Jawaharlal Nehru University, New Delhi in 2015. He also received his Master of Computer Application and Master of Technology degree in Computer Science and Technology from Jawaharlal Nehru University in 2009 and 2011 respectively. His research interests include support vector machines, extreme learning machines and other data mining techniques.