Paper ID: E&TC04

BACK-PROPAGATION NEURAL NETWORK AND SUPPORT VECTOR MACHINE COMPARISON FOR CLASSIFICATION OF STATIC HAND GESTURES

Mrs.Anjali R.Patil
PhD Scholar,
Shivaji University,
Kolhapur (Maharashtra)

Dr. Mrs. S. Subbaraman
Professor, Electronics Department
Walchand College of Engineering, Sangli.
(Maharashtra)

Abstract: Hand gesture recognition is an essential task as its application ranges from human-machine interface to sign language recognition. The classification accuracy of static gestures depends on the technique used to extract the features as well as the classifier used in the system. Artificial Neural Network (ANN) and Support Vector Machine (SVM) are the powerful mechanisms used for pattern recognition. This paper briefs the basic theory of ANN and SVM, and also provides the comparison of the Back propagation Neural Network and Support Vector Machine used as classifier for the static hand gestures.

Keywords: Static hand gesture, Back-propagation Neural Network, Maximum margin, Support Vector Machine, Classification accuracy.

I. INTRODUCTION

Hand gesture recognition plays an important role in wide area covering the applications from virtual reality to sign language recognition. Hand gestures are broadly classified as Glove based and vision based Gesture Recognition. In Glove based approaches users have to wear cumbersome wires which may hinder the ease and naturalness with which the user interact with the computers or machines. The awkwardness in using gloves and other devices can be overcome by using vision based systems that means video based interactive systems. This technique uses cameras and computer vision techniques to understand the gestures in a simple way. This approach is called as vision based approach. Vision based hand gesture systems have been broadly used by many researchers as it does not require any additional hardware other than camera. That is how the vision based techniques are suitable for computing and recent emerging applications. Methods for vision based hand gesture recognition are divided into two categories namely 3D model based methods and appearance model based methods [1] [2]. 3D model may exactly describe hand movement and its shape, but most of them are computationally expensive. In this paper, we focus on appearance model based methods for recognition of hand posture.

Many researchers, as can be seen from literature review, have worked on appearance based methods. Freeman et. al. recognized gestures which was further used for controlling the television using normalized correlation [3]. The method proposed in that paper works satisfactorily only for the static conditions. However it does not work for the variations in the pose and the background of a person performing hand gesture. Weng et. al. proposed a hand tracking and sign recognition method using appearance based method [4]. Even though its classification accuracy was good, the performance was not good in the real-time environment. Just et. al. used modified census transform for hand gesture classification with good results[5]. J. Triesch et.al. used elastic graphs to represent hands in different hand postures with local jets of Gabor filters [6]. This helped to locate hands without separate segmentation mechanism. However, the learning of a classifier was carried out using only a small set of image samples resulting into limited generalization of a method. The exhaustive literature review on vision based gesture recognition states that the performance of the system highly depends on the pose, variation in the background, illumination conditions. Some of the works carried out to overcome the above mentioned limitations are presented below.

Lars et.al. used scale-space color features for recognition of hand gestures [7]. In their approach they developed feature detection under uniform background and reported as real time with user independence and invariant to camera movement. The other researchers have dealt with gesture recognition problems using soft computing approaches to statistical models based on Hidden Markov Model (HMM) [15], and Finite State Machine (FSM) based models [16]. Soft computing tools generally include ANN [17] [18], fuzzy Logic sets [19] and Genetic Algorithms GAs [20]. Minimum Enclosing Ball-SVM (MEB-SVM) algorithm was implemented on large scale data classification by Yu Ren et.al. wherein mean shift and Fourier descriptor were implemented to develop an efficient and fast system. [21].

This paper presents the context of classification through SVM and NN approach, and provides an easy way to relate each class. The rest of this paper is organized as follows: Section 2 presents discussion on Neural Network followed by brief theory on Support Vector Machine in Section 3. Detailed results and comparisons are presented in Section 4. Conclusions are given in Section 5.

II. NEURAL NETWORK

The aim of this section is to outline basic theory of Artificial Neural Networks (ANN), which in this paper is used for classification and further for recognition. It is being used now-a-days for extensive applications such as pattern recognition and classification, data compression and optimization [16]. The most widely applied neural network algorithm in classification remains the feed forward back propagation algorithm.

Emulation of the biological neurons are done using individual nodes in a neural network which takes the input and performs computations on the input and passes the output to the other neuron based on some measures. The output of each node is associated with some activation function and also weight values are associated with the node in the network and, these values constrain how input data (In this case hand postures images) related to output data (e.g., Specific Hand postures). Weight biases are determined by the iterative flow of training data through the network (i.e., weight values are established during a training phase in which the network learns how to identify a particular class by typical input data characteristics) [17].

Most popularly used training algorithm is Back-propagation which comes under supervised learning technique used to train the neural network. The term is an abbreviation for "backward propagation of errors". Back-propagation takes the transfer function used by the artificial neurons or nodes are differentiable. Gradient of the error of the network with respect to the network's modifiable weights is calculated using Back propagation algorithm. This gradient is almost always then used in a simple stochastic gradient descent algorithm to find weights that minimize the error. The algorithm mostly contains the calculation of the gradient and its use in stochastic gradient descent. The beauty of this algorithm is that it normally allows quick convergence on satisfactory local minima for error. Non-linear activation functions that are commonly used include the logistic function, the Softmax function, the Gaussian functions and the sigmoid function.

The general model for neural network discussed above is as shown below in Fig. 1:

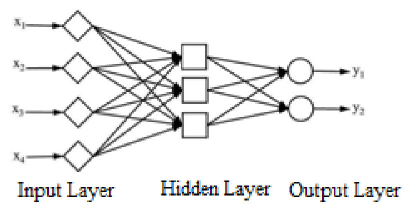

Fig. 1. General model of NN

Neurons are the fundamental processing unit which is the base element in neural network. The very basic processes involved in the neuron model can be identified as follows:

$$\text{Neuron's output} = f\left(\sum_{j=1}^{p} x_i w_j\right) \quad (1)$$

w_j -synaptic weights
x_i - Neuron inputs
f ( ) -activation function

Each of the input signals x_j is multiplied by the corresponding synaptic weight $w_{j..}$ Output of each neuron is obtained by applying a nonlinear function, called the activation function to the sum of the weighted input signals.

Mean Square Error (MSE) is calculated using following expression:

$$MSE = \frac{1}{N}\sum_{t=1}^{N}(F(x_t;W) - y_t)^2 \quad (2)$$

where, $(F(x_t;W)$ is the actual output of the network and y_t is the target output. If the MSE is not below a certain pre-determined value, the weights are updated as per follows:

$$\Delta w_i^j = -\eta \cdot \frac{\partial E}{\partial w_i^j}(W) \quad (3)$$

$$w_i^{j,new} = w_i^j + \Delta w_i^j \quad (4)$$

Paper ID: E&TC04

Here, η (0<η< 1) is the learning rate which is a scaling factor indicating the strength of the connection weights to be adjusted for a given error. Higher value of η can be used to speed up the learning process, however with a risk of skipping the optimum weights. It is proved that the algorithm will converge if training data is linearly separable and η sufficiently small.

III. Support Vector Machine

Support Vector Machine (SVM), proposed in 1990's by Vapnik is based on the linear discriminate functions used for classification and regression [16]. The fundamental concept of the SVMs is to build a hyper plane as the decision making plane, which is used to separate the binary classes with positive (+1) and negative (-1) with the largest margin, which is computed by finding the minimum Vapnik Chervonenkis (VC) dimension of SVM. In case of the binary classification problem the feature extraction is required to be initially performed. In the proposed method the 20 dimensional Fourier Descriptors are used for training SVM. Let us take the notations for the training data $xj \in \Re^d$ with a label $yj = \{-1,+1\}$, for all the training data j = 1, 2, 3, … l , where l is the number of data, and d is the dimension of the classification problem. When the two classes are linearly separable in $\Re^d$, separating hyper plane which gives the smallest generalization error among the infinite number of possible hyper-planes is required to be computed. Such an optimal hyper-plane gives the maximum margin between the two classes. The margin is calculated as the sum of the distances from the hyper-plane to the closest data points of each of the two classes. These closest data points are called Support Vectors (SVs). The middle line on Fig. 1 represents the optimal separating hyper-plane. Hyper-plane is represented with a direction vector w (normal to the hyper-plane) and an offset vector b that satisfies the equation

$$w^t.x + b = 0 \tag{5}$$

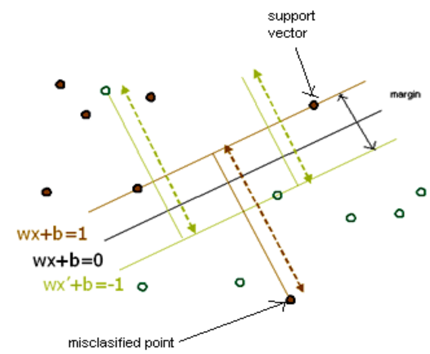

Fig. 2. Linear separating hyper-plane and Support vectors. The separation problem is to determine the hyper-plane such that $x_i + b \geq 1$ for positive examples and $wx_i + b \leq 1$ for negative examples.

In case of binary classification in the two-dimensional case, there are support lines, instead of planes, and the decision boundary also is a line as shown in the figure. The distance from any point x on the line $w_x + b = 1$ to the decision boundary is given by

$$\frac{f(x)}{||x||} = \frac{1}{||w||} \tag{6}$$

In the above, note that f(x) = 1 for any x on the line $wx_i + b = 1$. Similarly, the distance from a point on the line $wx_i + b = -1$ to the decision boundary is given by $\frac{-1}{||w||}$. So, the distance between the two supporting lines, or the margin, is $\frac{2}{||w||}$. Margin is maximized by maximizing $\frac{2}{||w||}$ or minimizing $\frac{||w||}{2}$ or for the sake of simplicity as in calculus, $\frac{||w||^2}{2}$

$$\min_{w,b} \Phi(w) = \frac{||w||^2}{2} \tag{7}$$

The optimal hyper-plane is found by minimizing (6) under the constraint (7) to correctly classify the training data.

$$y_i = (w.x_i + b) - 1 \geq 0, \forall i \tag{8}$$

Implementing linear SVM described above is a Quadratic Programming (QP) problem for which standard techniques like Lagrange Multipliers can be used.

$$f_\lambda : \Re^n \rightarrow \{-1,+1\} \tag{9}$$

The set of functions f_λ could be a set of Radial Basis Functions or a multi-layer neural network.

IV. Experimental set-up and Results

For performing the hand posture classification shape based features were used and the performances of the two said classifiers were compared. The comparison was done based on various performance parameters like Classification Rate (CR) or Classification Accuracy (CA), False Alarm Rate (FAR), False Rejection Rate (FRR).

In the training database total 6 postures A, B, C, Point, Five and V from Sebastian Marcel Database [19] have been used. Sample postures are given in Fig. 3:

Fig. 3. Six postures ‘a’, ‘b’, ‘c’, ‘five’, ‘point’, ‘v’ from Sebastian Marcel static hand postures database[19].

300 samples were used in the training phase (50 samples of each posture) and in testing 100 samples were used. Out of 100 samples in the testing, 50 were used for calculation of False Acceptance Rate (FAR) and remaining 50 were used for finding False Rejection Rate (FRR). The size of posture images selected was 100 x100. The algorithm first applies image pre-processing techniques on the images in order to cancel background and noise effects. The algorithm used for extracting the features is as follows:

1. Input static hand gesture color image.
2. Preprocess the input image using following steps:
   i) Convert RGB image to YUV color space.
   ii) Extract the skin color by applying thresholding.
   iii) Noise removal by applying median filter.
   iv) Convert the color image to black and white image with gesture representing binary as 1 and background as 0.
3. Noise (eg. Spikes) is removed using morphological filters.
4. Crop, and resize the extracted hand gesture to a size of 40 x 40 pixels in order to have the dimension of the Fourier descriptor uniform for all the images in the training database.
5. Extract the boundary pixel coordinates of the gesture image.
6. Represent the boundary in the complex plane where the column-coordinate is the real part and the row-coordinate the imaginary part.
7. Calculate Fourier transform and form the feature vector of the 20 dimensional

The features that are calculated using above algorithm are further used to classify the gestures using NN and SVM. Fig. 4 below shows the result of the algorithm which was used to extract the features of posture ‘A’. Fig. 5 shows the boundary detection of the different hand postures used in this paper.

Fig. 4. Steps in Feature extraction of sample posture ‘A’.

Fig. 5. Boundary Detection results of the proposed system for posture Five, B, C, V, and Point.

In proposed system, multilayer feed forward neural network has been used for classification. For learning the network, Back propagation algorithm is used. The activation function used is “Sigmoid”.

The second classifier used to classify six hand gestures (and not only two) was the combination of binary SVMs in a Classifier Tournament structure. For every pair of classes, a SVM classifier was trained that classifies between them. After new image features were computed in the preprocessing stage, the features were entered into all the binary classifiers. Every binary classifier outputs a class number giving a specific hand posture. Voting is taken and maximum number was the final class of posture. Detailed discussion on different multiclass SVM can be found at [20].

Paper ID: E&TC04

A. Performance parameters:

1) False acceptance rate (FAR): is a measure of probability how a recognition system will accept an incorrect input as a positive match. A system's FAR typically is stated as the ratio of the number of false acceptances divided by the number of identification attempts.

$$FAR(\%) = \frac{FalseMatches}{impostorAttempts} \quad (17)$$

2) **False Rejection rate (FRR)**: By contrast the FRR is a measure of probability a system will incorrectly reject an input as a negative match. A system's FRR typically is stated as the ratio of the number of false rejections divided by the number of identification attempts.

$$FRR(\%) = \frac{FalseNonMatch}{EnroleeAttempt} \quad (18)$$

3) Classification accuracy: It is defined as the rate of positive decisions in the total number of decisions for each stage of the procedure for instance positive match for all the test image will lead to 100% Recognition Accuracy. The action or process of classifying the correct postures from the test database is Recognition rate.

$$CA(\%) = \frac{PositiveMatch}{EnroleeAttempt} \quad (19)$$

Table 1 and 2 shows the confusion matrix obtained from the SVM and ANN classifier. Total 100 images of each posture are used for testing the machines. Table 3 represents the FAR, FRR and CA for the six postures.

TABLE 2. CONFUSION MATRIX (ANN AS CLASSIFIER)

Postures	A	B	C	POINT	FIVE	V
A	90	-	6	-	-	4
B	1	93	-	6	-	-
C	-	-	93	7	-	
POINT	-	-	2	91	-	7
FIVE	1	-	3	-	92	4
V	3	3	-	-	-	94

TABLE 2. CONFUSION MATRIX (SVM AS CLASSIFIER)

Postures	A	B	C	POINT	FIVE	V
A	84	-	7	1	-	8
B	3	89	-	6	2	-
C	2	1	91	5	-	1
POINT	7	-	2	84	-	7
FIVE	4	2	-	-	92	2
V	4	5	1	-	-	90

TABLE 3. PERFORMANCE OF NN AND SVM CLASSIFIERS:

Classifier	Static Hand Gesture	No. of test images	FAR %	FRR %	CA %
BP-NN	A	50	5	10	90
	B	50	3	7	93
	C	50	11	7	93
	Point	50	13	9	91
	V	50	15	8	92
	Five	50	0	6	94
Average Classification Accuracy					92.16
SVM	A	50	7.2	28	84
	B	50	2.8	9	89
	C	50	4.8	22	91
	Point	50	2.6	29	84
	V	50	9.2	29	92
	Five	50	00	36	90
Average Classification Accuracy					88.33

V. CONCLUSION AND FUTURE SCOPE

The comparative study of NN and SVM classifiers for static hand gesture classification and recognition for the same feature vectors is reported with reference to CA, FAR and FRR.

Overall, results show the performance of these two different classifiers are close and comparable. Average classification accuracy for ANN is 92.16 % and that of using SVM is 88.33%. Database that has been used in this paper consist of the posture images having different view and illumination. The accuracy can be further increased by selecting the robust features which will be invariant to view and illumination.

Future work will include extending the developed method to recognize dynamic Hand Gestures with video based and interactive system. This system can be implemented in various environments such as PC, mobile equipments and embedded etc.

References:

[1]. Vladimir I. et.al. 'Visual Interpretation of Hand Gestures for Human-Computer Interaction: A Review', IEEE transactions on Pattern Analysis and Machine Intelligence, vol. 19, no. 7, July 1997.

[2]. Sushmita Mitra, Tinku Acharya, 'Gesture Recognition: A Survey' IEEE Transactions on Systems, Man, and Cybernetics—Part C:

Applications and Reviews, Vol. 37, No. 3, MAY 2007.

[3]. W. T. Freeman and C. Weissman, "Television control by hand gestures," in Proceedings of International Workshop on Automatic Face and Gesture Recognition. Zurich, Switzerland, June 1995, pp. 197–183

[4]. Y. Cui and J. Weng, "View-based hand segmentation and sequence recognition with complex backgrounds," in proceedings of 13th ICPR. Vienna, Austria, Aug. 1996, vol. 3, pp. 617–621.

[5]. Just A., Rodriguez Y., and Marcel S., "Hand posture classification and recognition using the modified census transform," in Proceedings of Int. Conf. on Automatic Face and Gesture Recognition. Southampton, UK, Apr. 2006, pp. 351–356.

[6]. J. Triesch and C. von der Malsburg. ―A System for Person-Independent Hand Posture Recognition against Complex Background ,IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol.23, No. 12, December 2001.

[7]. Lars Bretzner, Ivan Laptev, and Tony Lindeberg, "Hand gesture recognition using multi-scale colour features, hierarchical models and particle filtering," in Proceedings of Int. Conf. on Automatic Face and Gesture Recognition. Washington D.C.,May 2002, pp. 423–428.

[8]. Mahmoud Elmezain, Ayoub Al-Hamadi, Jorg Appenrodt, and Bernd Michaelis. (2009). "A Hidden Markov Model-Based Isolated and Meaningful Hand Gesture Recognition", International Journal of Electrical and Electronics Engineering, pp. 156-163.

[9]. Verma R., Dev A. (2009, Dec.). "Vision based hand gesture recognition using finite state machines and fuzzy logic", IEEE International Conference on Ultra Modern Telecommunications & Workshops, (ICUMT '09), pp. 1-6, Petersburg, doi:10.1109/ICUMT.2009.5345425.

[10]. Tin Hninn H. Maung. (2009). "Real-Time Hand Tracking and Gesture Recognition System Using Neural Networks", World Academy of Science, Engineering and Technology 50, pp. 466- 470.

[11]. Manar Maraqa, Raed Abu-Zaiter. "Recognition of Arabic Sign Language (ArSL) Using Recurrent Neural Networks," IEEE First International Conference on the Applications of Digital Information and Web Technologies, (ICADIWT 2008), pp. 478-48. Available: doi: 10.1109/ICADIWT.2008.4664396

[12]. Xingyan Li. (2003). "Gesture Recognition Based on Fuzzy C-Means Clustering Algorithm", Department of Computer Science. The University of Tennessee Knoxville.

[13]. Aviad Barzilai, Adi Fuchs, Michael Kolomenkin, "Hand Gesture Classification Using Depth Information", CGM Laboratory undergraduate project, Electrical Engineering Dept., Technion – IIT, 2009

[14]. Yu Ren Fengming Zhang,'Hand Gesture Recognition Based on MEB-SVM',2009 International Conference on Embedded Software

[15]. Nasser H. Dardas and Nicolas D. Georganas, "Real-Time Hand Gesture Detection and Recognition Using Bag-of-Features and Support Vector Machine Techniques". IEEE Transactions on Instrumentation and Measurement, vol.60, no.11, November 2011.

[16]. Neural Networks for Classification: A Survey Guoqiang Peter Zhang International Journal of Signal Processing, Image Processing and Pattern Recognition Vol. 2, No. 4, December, 2009.

[17]. E. Stergiopoulou, N. Papamarkos, 'Hand Gesture recognition using a neural network shape fitting technique',ELSEVIER Engineering Applications of Artificial Intelligence 22 (2009) 1141–1158

[18]. Takefumi Matsunaga and Oshita Masaki, "Recognition of Walking Motion Using Support Vector Machine", Proc. of ISICE 2007, pp. 337-342, 2007.

[19]. S. Marcel, "Hand posture recognition in a body-face centered space," in Proc. Conf. Human Factors Comput. Syst. (CHI), 1999, pp. 302Processing, 2005, vol. 2, pp. 954 - 957.

[20]. Kai-Bo Duan, S. Sathiya Keerthi, 'Which Is the Best Multiclass SVM Method? An Empirical Study', Springer Berlin, 2005.

[21]. Chih-Wei Hsu et.al. 'A Practical Guide to Support Vector Classification', Department of Computer Science National Taiwan University, Taipei 106, Taiwan April 15, 2010