 ISSN: 2321-8169
81 – 84

Comparative Study between MLP and SVM as a Classifier for Voice Recognition

Y. M. Pharkade
Degree College of Physical Education
HVPM
Amravati (M.S.), India

Dr. G. D. Agrahari
Brijlal Biyani science College
Amravati (M.S.), India
gddagrahari@gmail.com
yogifr123@gmail.com

Dr. D. S. Dhote
Brijlal Biyani science College
Amravati (M.S.), India
dsdhote@rediffmail.com

Abstract— using the physical and behavioral characteristics, the person may be identified in a crowd. Usually, the biometric systems are design and develop by taking the inputs features from the physical characteristics. A neural network can be developed for the identification of a person. Here, in this, the features are extracted from the voice samples of a group of peoples. The Multilayer Perceptron and Support Vector Machine classifiers are used to classify and recognition of voice samples. From the result of the experiment, it is found that Support Vector Machine network having the best classification accuracy and the network becomes the more suitable. It is found that the SVM network is fast and having more accuracy to identify the person from his/her vowels.

Keywords- *Artificial Neural Network, MLP, SVM.*

*****

I. INTRODUCTION

The study is based on the voice features of a person by using which he/her can automatically identify and using the voice command, the person can easily interact with the machine. By adapting some method of human communication, it is expected that human interaction with machines is also becoming more natural and can even facilitate the people with special needs such as persons with disabilities [1]. Some of the interaction method includes eye, pronunciation, touch of the hand etc.[2]. Interaction between the user and the devices usually takes place via advanced 'natural' user interaction techniques involving human-speech [3].Voice recognition have very wide range of applications in security systems and Robotics. Voice Recognition system[4] is used to recognize a person by using his or her vowel. voice recognition is a more personalized form of control, since it can be adapted and customized to a particular speaker's voice [5].

II. LITERATURE REVIEW

According to Ananth Sankar and Richard J. Mammone of CAIP Center and Dept. of Electrical Engineering, Rutgers University, vowel recognition is a difficult pattern recognition problem in the Voice recognition system. Recently there has been much research using Multi-Layer Perceptrons (MLP) and Decision Trees for this task. Voice recognition experiments on monosyllabic vowels and vowels in words. This paper was presented by Shimodaira, H.; Klimura,M at 9th International Conference on Pattern Recognition, in 1988.In that they presented a vowel recognition technique that deals with input vowels as one sequence and recognizes them using binary relationships between data is investigated. Voice recognition experiments on monosyllabic vowels and vowels in words showed that recognition rates are 99.1% and 93.2%, respectively. These results are about 2% better than those obtained using a statistical matching method.. Badran and Selim, H.of Assiut University were presented a paper in 5th International Conference on Signal Processing in year 2000. They were worked on the Voice recognition systems that attempt to recognize a person by his/her voice through measurements of the specifically individual characteristics arising in the person's voice.

III. DETAILS OF EXPERIMENT

The voice recognition system consisting of neural network is a well trained network having sufficient amount of data for training, cross- validation and testing. The benchmark data sets is get available from Principie. The data is stored in the column vectors of excel-sheet and then it is supplied to neural-network. The accompanying data file, "vowelcontext.data", consists of a three dimensional array: voweldata [speaker, vowel, input]. There are fifteen individual speakers. The names of these speakers are Andrew, Bill, David, Mark, Jo, Kate, Penny, Rose, Mike, Nick, Rich, Tim, Sarah, Sue, Wendy. These 15 speakers are assigned with a code number. Each individual speaker repeats each vowel six times. So there are total $15 \times 6 = 90$ voice samples. This data is indexed by integers 0-89. Total eleven vowels are used for train and test the NN. These vowels are hid, hId, hEd, hAd, hYd, had, hOd, hod, hUd, hud, hed and indexed by integers 0-10.

Each person utters each vowel six times, so total 66 rows (11×6) of data for each person. For each utterance, there are ten features extracted from these input values i.e. Feature 1, Feature 2,….Feature 10.These features are floating-point values. Also there is, Sex_column to give one additional feature to the network. The speech signals were

 ISSN: 2321-8169
81 – 84

low pass filtered at 4.7 kHz and then digitized to 12 bits with a 10 kHz sampling rate. As the "vowel-context .data" data file consists of space separated values, the file can directly import in NeuroSolution. After saving the data file, tagging must be done on input and desired output column and select the percentage of data for training, cross validation and testing are as given in Table 1.

TABLE 1 PERCENTAGE OF DATA FOR EXPERIMENT

Total Dataset=990	**Data Partition**	**Number of Data samples**
Training	60%	594
Cross Validation	20%	198
Testing	20%	198

The parameters such as MSE, Final MSE and visual inspection of desired output are to be taken to assess the performance of NN. The Network has been trained at least 3 times, starting from different random initial weights so as to avoid local minima. Neurosolution is specially used for obtaining results. The comparative study for the voice recognition using MLP and SVM has been carried out.

IV. RESULTS OF EXPERIMENT

The extracted features from voice samples are taken as input and a particular person having its voice is the targeted output. The Neurosolution software require these features as input and gives the desired output. The parameters such as Processing elements, transfer function, learning rule, step size and momentum of hidden layer and output layer of MLPs were tested with maximum epoch 5000 and for the 3 runs.

A. Selection of Transfer Function and Learning Rule for Hidden Layer:

Various learning rules such as step, momentum, CG, LM, QP and DBD are used for training and best performance parameters are observed. Results are observed for various Transfer Functions like TanhAxon, SigmoidAxon, Linear TanhAxon, Linear SigmoidAxon, SoftMaxAxon, BiasAxon, LinearAxon and Axon only as shown in the Table 2.

TABLE 2 TRANSFER FUNCTIONS AND LEARNING RULES IN HIDDEN LAYER

Transfer function	**Learning rule**	**MSE**
TanhAxon	Step	0.020361704
TanhAxon	Mom	0.013935919
SigmoidAxon	CG	0.002886206
SigmoidAxon	**DBD**	**0.002474252**
Linear Axon	Step	0.1008
Linear Axon	Mom	0.076990809

From Table 2, it is inferred that the optimal values are obtained for Transfer function –SigmoidAxon and Learning Rule – DBD in the hidden layer. The values of MSE= **0.002474252** which are optimal values as compared to the results obtained for other transfer function.

TABLE 3 HIDDEN LAYER PARAMETERS

In the Hidden Layer , Transfer Function- SigmoidAxon and Learning Rule –DBD **In the Output Layer, Learning Rule- DBD** **No. of Epochs=5000**	
Transfer Function	**MSE**
TanhAxon	0.0607995
SigmoidAxon	**0.0032995**
LinearAxon	0.0880341

B. Selection of Step Size and Momentum Rate for Output Layer and Hidden Layer:

MLP is also called Back Propagation. During training, iteration of back propagation NN, the weights at the output layer are modified and by proceeding backwards through the hidden layers one by one until it reach the input layer. It is the method of proceeding backwards. Therefore weights and biases are back propagating to the previous layer.

The network was trained for 5 times by varying Step Size and Momentum Rate individually. It was noticed from table 4 and 5 that the optimal values are obtained at **Step Size = 1.0** and **Momentum Rate = 0.6**

TABLE 4 STEP SIZE VS DIFFERENT PARAMETERS FOR OUTPUT LAYER

Step Size	0.6	0.7	0.8	0.9	1.0
MSE	0.0035	0.0036	0.0036	0.0038	**0.0031**

 ISSN: 2321-8169
81 – 84

TABLE 5 MOMENTUM RATE VS DIFFERENT PARAMETERS FOR OUTPUT LAYER

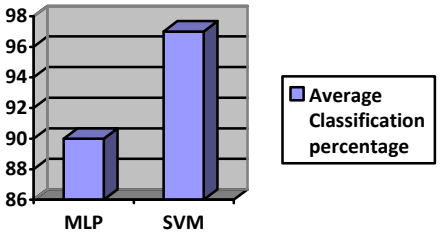
Momentum Rate	**0.6**	**0.7**	**0.8**	**0.9**	**1.0**
MSE	**0.003**	0.0034	0.0040	0.005	0.025

And now these values of Step Size = **1.0** and Momentum Rate = **0.6** are fixed for the Hidden and Output Layer for developing the MLP Network. The below Table no. 6 summarizes the result for this classification of persons with his or her vowels, having **average classification Percentage= 90 %**

TABLE 6 SUMMARY OF THE RESULT USING MLP

Performance ➡ Name of person ⬇	**MSE**	**NMSE**	**r**	**Percent Correct**
Bill	0.01425	0.26412	0.86450	97.059
Sue	0.00592	0.12330	0.95676	100
Nick	0.01297	0.21140	0.89582	97.4359
Mark	0.03199	0.43063	0.77731	72.9167
Wendy	0.01174	0.23733	0.89244	100
Mike	0.02761	0.42966	0.75915	87.8049
Rose	0.00342	0.06535	0.98772	100
Penny	0.0089	0.16003	0.9266	100
Tim	0.02499	0.4172	0.81597	76.3158
Kate	0.01203	0.1621	0.93309	89.5834
Sarah	0.03892	0.5922	0.7083	57.14286
Rich	0.01405	0.21851	0.89977	95.122
David	0.00488	0.0778	0.97442	100
Jo	0.01726	0.2415	0.8893	86.957
Andrew	0.00586	0.0789	0.9773	97.917

C. SVM used as a Classifier:



The Support Vector Machine (SVM) is a new kind of classifier that is motivated by two concepts. First, transforming data into a high dimensional space can transform complex problems into simpler problems. Second, SVMs are motivated by the concept of training and using only those inputs that are near the decision surface since they provide the most information about the classification. In addition to the normal voice, an abnormal sound, such as screaming and glass broken, were discriminated from normal sounds where Gaussian Mixture Model (GMM) and Support Vector Machine (SVM) were commonly used for recognitions [6]-[8].

Table no. 7 shows the result for the classification of Speakers, having **average classification Percentage= 97 %**

TABLE 7 SUMMARY OF THE RESULT USING SVM

Performance ➡ Speaker ⬇	**MSE**	**NMSE**	**r**	**Percent Correct**
Bill	0.00335	0.062	0.9869	100
Sue	0.00290	0.06067	0.99233	100
Nick	0.00681	0.11094	0.9646	94.872
Mark	0.011	0.1478	0.9448	91.667
Wendy	0.00489	0.987	0.97627	100
Mike	0.0144	0.2238	0.9061	85.366
Rose	0.00283	0.0540	0.9962	100
Penny	0.0032	0.0581	0.9892	100
Tim	0.006	0.0989	0.9696	97.37
Kate	0.0035	0.0466	0.991	100
Sarah	0.0049	0.075	0.9805	97.62
Rich	0.0031	0.0485	0.9893	100
David	0.0028	0.0443	0.994	100
Jo	0.00831	0.1164	0.959	93.48

 ISSN: 2321-8169
81 – 84

Andrew	0.00296	0.0398	0.996	100

V. CONCLUSION

The result of both the classifier i. e. MLP and SVM, from table no. 5 and 6 shows that, for MLP the Average classification percentage is 90% while for the SVM, the average classification percentage is 97%.

Hence the network model using Support Machine vector (SVM) is found to be perfect which gives 97% accuracy for identifying the voice from his/her vowels and it gives the best result compared with Multilayer Perceptron.

ACKNOWLEDGMENT

We are using this opportunity to express our gratitude to everyone who supported us throughout the research work. We are thankful for their aspiring guidance and friendly advice during the work. We are sincerely grateful to them for sharing their truthful views on a number of issues related to the research work.

REFERENCES

[1] Muslim Sidiq, Tjokorda Agung Budi W, Siti Sa'adah, "Design and Implementation of Voice Command Using MFCC and HMMs Method"

[2] A. K. Andrey Ronzhin, "Assistive Multimodal System Based On Speech Recognition And Head Tracking," dalam SPIIRAS, St. Petersburg, Russia.

[3] T. Koskela and K. V¨a¨an¨anen-Vainio-Mattila, "Evolution towards smart home environments: empirical evaluation of three user interfaces,"Personal and Ubiquitous Computing, vol. 8, no. 3-4, pp. 234–240, 2004.

[4] Khalid Saeed, "A Note on Biometrics and Voice Print: Voice-Signal Feature Selection and Extraction - A Burg-Toeplitz Approach", in Proc. of IEEE, Scientific Workshop Signal Processing'2006.

[5] Yash Mittal, Paridhi Toshniwal, Sonal Sharma "A Voice-Controlled Multi-Functional Smart Home Automation System", IEEE INDICON 2015

[6] Jianzhao Qin, Jun Cheng, Xinyu Wu, and Yangsheng Xu, "A learning based approach to audio surveillance in household environment," *International Journal of information Acquisition*, vol. 3, no. 3, pp. 1-7, 2006.

[7] Huy Dat Tran, and Haizhou Li, "Sound event recognition with probabilistic distances SVM," *IEEE Transaction on Audio, Speech, and Language Processing*, pp. 1556-1568, vol. 19, no. 6, August 2011.

[8] M. A. Sehili, D. Istrate, B. Dorizzi, and J. Boudy, "Daily sound recognition using a combination of GMM and SVM for home automation," *Proc. of 20th European Signal Processing Conference*, pp.1673-1677, 2012.