

Comparative Performance Analysis of SVM Speaker Verification System using Confusion Matrix

Piyush Mishra¹, Piyush Lotia²

^{1,2} Department of Electronics & Telecommunication Engineering, SSCET, Bhilai, Chhattisgarh, India

Abstract: In Speaker verification task, it is necessary to calculate the performance of the speaker verification system; there are many systems available for the speaker verification task which uses the different type of modeling schemes like generative modeling and discriminative modeling. We are using discriminative modeling with the help of Support vector machine for speaker verification task. We propose the use of confusion matrix for the performance calculation of support vector machine in contrast with the parameters like accuracy and precision. Speaker verification includes training of feature vector and then the classification of trained feature vector. Every kernel of support vector machine gives the different performance for the classification task so with the help of this confusion matrix approach we can also compare the different kernel performances. The parameters of performance make us capable for selecting the best kernel for our data sets.

Keywords: Classifier, kernel, generative and discriminative model, Confusion matrix, SVM, Verification, Support Vector.

1. Introduction

In last decade the discriminative modeling approach becomes very successful in the classification task. The support vector machine uses the discriminative modeling approach and creates the data sets of different class. These data sets are called support vectors, support vector are classified with the help of linear and non linear transformation. The linear transformation used if the support vectors are easily separable and for non separable vectors the non linear transformation is used with the help of kernels like linear, quadratic, polynomial, radix base function etc. Confusion matrix contains the information about the classified support vector machine.

Confusion matrix calculates the accuracy, precision, recall, F-measure with the help of True positive rate (TPR), True negative rate (TNR), false positive rate (FPR), and (FNR). The term F-measure is also called F1 score is the measure of the performance test accuracy, term recall also known as sensitivity is the fraction of relevant classes that are retrieved. We are using the voice of 3 speakers with the training ratio of 50 percent. The number of coefficient of each speaker voice is selected in such a manner that dimension of the SVM structure does not exceed.

2. Training of Support Vectors

The algorithm starts with training the data with support vector machine, for each of the voice samples the sampling frequency and the number of bits used for encoding is determined. The sampling frequency and number of bits for encoding is chosen according to the structure of support vector machine to maintain the dimension. Since the temporal domain characteristics are complicated to determine so the spatial (frequency) domain characterization is done with the help of power spectral density. Each speaker voice is then converted into vectors called feature vector. Since the feature vector obtained is two dimensional we have

to convert it into the three dimensional support vector. The three dimensional vector creates two training sets called training set 1 training set 2. Training set 1 and 2 are called feature vector of class 1 and 2 respectively.

The three dimensional support vectors are trained with the help of support vector machine. The structure of support vector machine is decided by the number of coefficients, total number of segments, segment duration and overlapping percentage. The overlapping is necessary to extract each of the coefficient more accurately from the power spectral density.

If the data is linearly separable then it can easily classify into two classes but if the data set or vectors are not separable we have to use kernels of support vector machine. Each and every kernel of support vector machine has its own transformed domain where the non linear data can easily be classified with the help of binary classifier.

3. Non Linear Classification with Kernel

The support vector machine has:

- 1) Polynomial
- 2) quadratic
- 3) Radial Basis Function
- 4) Multilayer perception many types of kernels

Suppose x and y are the vector of two classes and they belongs to non linear class, consider the function denoted $k(x,y)$ called kernel function transforms the data into the domain where it can be easily separable. Each kernel used in support vector machine has the different domain of transformation shown below:

- a) Polynomial kernel: – the transformation function for polynomial kernel

$$K(x, x_i) = (1 + x_i^T x / c)^d \quad (1.1)$$

b) X_i represents the input data in the matrix form, X_i^T represents the transpose of the matrix X_i and d is the degree of the polynomial kernel.

c) Quadratic kernel: – it is the special class of polynomial kernel with $d=2$

d) Radial Basis Function: – it is abbreviated as RBF kernel, also called Gaussian Radial Basis Function denoted

$$K(x, x_i) = \exp(-\|x - x_i\|^2 / \sigma^2) \quad (1.2)$$

e) Radial basis function is related to Gaussian density function where $X - X_i$ represents the deviation from the mean value while σ^2 represents the variance.

f) Multilayer perception kernels: – It is also known as mlp kernel has the transformation based on hyperbolic tangential function

$$K(x, x_i) = \tanh(kx_i^T x + \theta) \quad (1.3)$$

X_i represent the input data in the matrix form, X_i^T represents the transpose of the matrix X_i and θ represent the offset value of the transformed domain. Consider the diagram how non linearly separable data converted into linearly separable data with the help of kernel function.

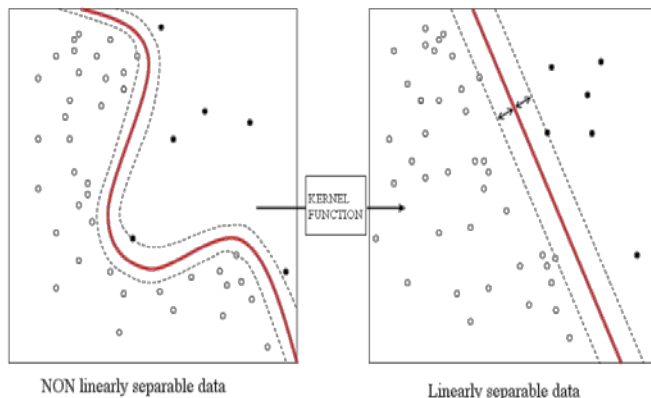


Figure 1: Separation of non linear data through kernel function

4. Performance Calculation

We are proposing the method for calculating the performance of the support vector machine with the help of confusion matrix. A confusion matrix consist the information about the actual data class and the predicted data class done by the classifier. The classifier used here is the support vector machine. The confusion matrix also known as contingency table.

		Predicted class	
		P	N
Actual class	P	TP	FP
	N	FN	TN

Figure 2: Basic structure of 2X2 confusion matrix

Let us define there are two class of data, the first class denoted P(positive) and the second class denoted N (negative). The four outcomes can be formulated as 2X2 confusion matrix,

True Positive (TP):– it is the number that indicates how many times actual class and predicted class both positive.

True Negative (TN):– it is the number that indicates how many times actual class and predicted class both negative

False Positive(FP) :– It is the number that indicates that how many times actual class is positive but predicted class is negative.

False negative (FN) :– It is the number that indicates that how many times actual class is positive but predicted class is negative.

With the help of all these terms TPR, TNR, FPR and FNR is calculated as follows

True positive ratio = TP/P

True Negative Ratio = TN/N

False Positive Ratio = FP/P

False negative Ratio = FN/N

The basic algorithm for calculation of all the confusion matrix elements is explained in Fig 3. Firstly the speaker voice is sampled and converted into feature vector called test data, the test data is applied to SVM where the domain transformation is done if the data set is not linearly separable. The output of SVM is called trained data which has two class. The two class data is applied to the confusion matrix elements to check the calculation depending on actual class and predicted class. If actual class and predicted class are equal then true class are detected and if they are not equal false class is detected.

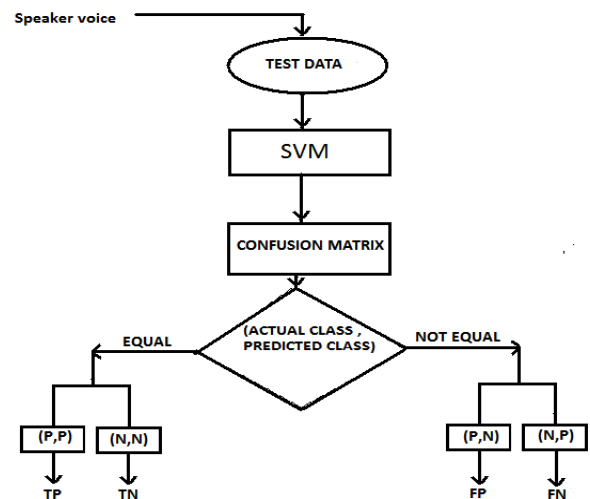


Figure 3: Algorithm of performance calculation

The performance parameters of confusion matrix are:

Accuracy - The accuracy is the proportion of the total number of predictions that were correct. It is Determined using the equation

$$\text{Accuracy} = (TP + TN)/(P + N);$$

Precision - Precision is the fraction of the class retrieved that is relevant to the true predictions

$$\text{Precision} = (TP)/(TP + FP);$$

Recall - Recall is the fraction of the classes that are relevant to the actual class that are successfully retrieved.

$$\text{Recall} = (TP)/(TP + FN);$$

F_measure – it is also called F1 score is the measure of the performance test accuracy.

$$F_measure = 2 * \text{Precision} * \text{Recall} / (\text{Precision} + \text{Recall});$$

5. Experiment Setup and Results

We performed the experiment of performance calculation on 3 speakers. To train the data each voice sample is converted into 50 segments, each segment has the duration of 2 seconds out of which 20 coefficients are taken from each segment with the overlapping percentage of 50. From each voice sample 50 percent is taken for the training so training ratio would be 0.5. Here we are using polynomial, quadratic, RBF and MLP. All these kernel functions have different accuracy, precision, recall and F-measure.

All the trained data through support vector machine is shown in figures, the polynomial kernel has better accuracy but it takes more time to train the data. The support vector trained through polynomial kernel is shown in Fig.4, it is also indicated that how the two class data is classified linearly.

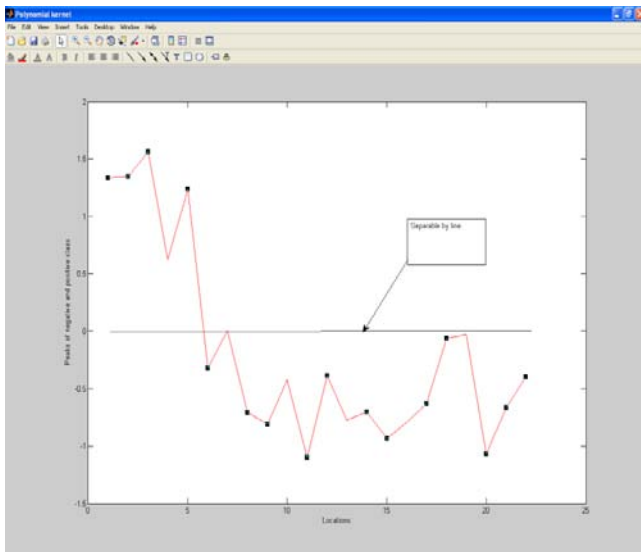


Figure 4: Trained support vectors through Polynomial kernel

The quadratic kernel is the special case of polynomial kernel with degree 2 it takes less training time. The trained support vector through quadratic kernel is shown in Fig .5 it is clearly seen that it has less margin as compared with the polynomial kernel,

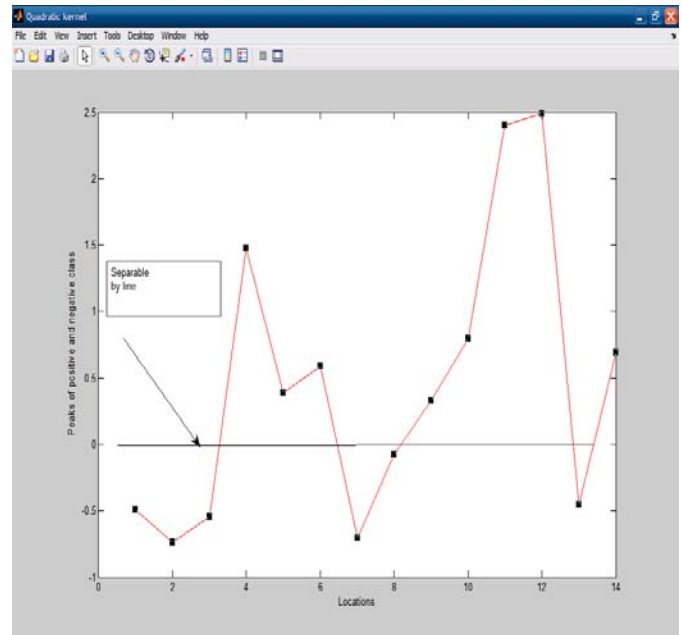


Figure 5: Trained support vectors through Quadratic kernel

RBF kernel provides better accuracy with larger margin between the classes but takes more matching time among all. It is indicated in Fig.6 larger the margin better the accuracy,

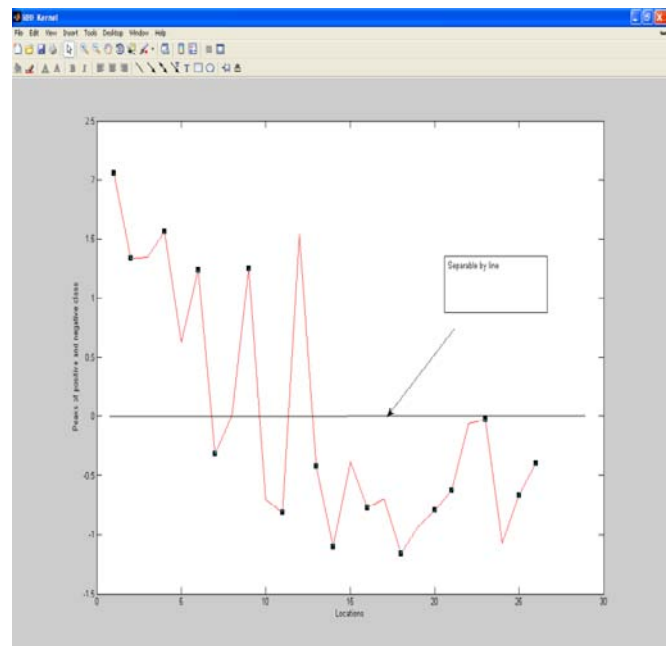


Figure 6: Trained support vectors through RBF kernel

It is indicated in Fig.7 MLP kernel has smallest margin among all so it has less accuracy with smallest value of F-measure

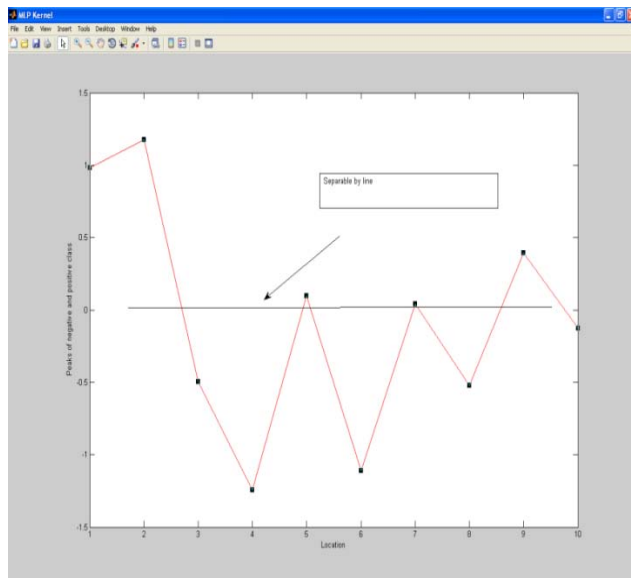


Figure 7: Trained support vectors through MLP kernel

The comparative result of all the kernel functions calculated by the two class data with the help of 2X2 confusion matrix is shown in the TABLE I

Table 1: Comparison of the different kernel function for the actual class and predicted class with the parameters true positive ratio, true negative ratio, false positive ratio and false negative ratio

Kernel function	TPR	TNR	FPR	FNR	Training time (sec)
Polynomial	0.90667	0.9553	0.046667	0.09333	5.2047
Quadratic	0.88	0.94	0.06	0.12	4.7963
RBF	0.89333	0.94667	0.05333	0.10667	4.9159
MLP	0.77333	0.88667	0.11333	0.22667	4.7769

By the parameters of TABLE I, the performance parameters of Support vector machine with the help of 2X2 confusion matrix is shown in TABLE II

Table 2: Comparison of the different kernel function for the actual class and predicted class with the parameters true positive ratio, true negative ratio, false positive ratio and false negative ratio

Kernel function	Accuracy	Precision	Recall	F measure	Matching time (sec)
Polynomial	0.93778	0.91578	0.90667	0.90612	0.00211
Quadratic	0.92	0.89003	0.88	0.87918	0.00208
RBF	0.92889	0.91919	0.89333	0.89468	0.00274
MLP	0.84889	0.79484	0.77333	0.76505	0.00213

6. Conclusion

We have presented the method for performance calculation for SVM speaker verification system. Our algorithm is suitable for every classifier with two or more than two number of classes. The experiment is carried out on 3 speakers with two level classification of positive and negative class. The support vector machines non linear classification contain polynomial, quadratic, RBF and MLP kernel function, so each kernel has different accuracy, precision, recall and F-measure. It is clearly seen by the comparison table that how the kernel is selected to get better performance parameters. The polynomial kernel provides

accuracy of 93.778 % and precision of 91.578%, but it takes training time of 5.204 sec which is the largest among all. It would be suggested that always maintain a tradeoff between training time and accuracy of kernel.

7. Future Scope

Future work will be based on performing the experiment on larger number of speakers and keeping the training time less with better accuracy and precision and the selection of better kernel for support vector machine.

References

- [1] V.Mohan Patro, Manas Ranjan Patra "Augmenting Weighted Average with Confusion Matrix to Enhance Classification Accuracy" paper published in TMLAI vol 2, issue 4 ISSN: 2054-7390
- [2] Najim Dehak (etal) "Support Vector Machines versus Fast Scoring in the Low-Dimensional Total Variability Space for Speaker Verification" Published at research and devolvement laboratory Paris France 2009.
- [3] William M. Campbell, Joseph P. Campbell, Terry P. Gleason, Douglas A. Reynolds "Speaker Verification Using Support Vector Machines and High-Level Features" iee transactions on audio, speech, and language processing, vol. 15, no. 7, September 2007.
- [4] Ing-Jr Ding, Chih-Ta Yen* and Da-Cheng Ou "A Method to Integrate GMM, SVM and DTW for Speaker Recognition" International Journal of Engineering and Technology Innovation, vol. 4, no. 1, 2014, pp. 38-47
- [5] W. M. Campbell, D. E. Sturim, and D. A. Reynolds, Senior "Support Vector Machines Using GMM Super vectors for Speaker Verification" iee signal processing letters, vol. 13, no. 5, may 2006
- [6] W. M. Campbell, J. P. Campbell, D. A. Reynolds, D. A. Jones, and T. R. Leek "high-level speaker verification with support vector machines" paper published at mit Lincoln laboratory Lexington, ma 02420, 2008
- [7] W. M. Campbell (etal) "svm based speaker verification using a gmm super vector kernel and nap variability compensation" paper published in MIT Lincoln Laboratory Lexington, MA 02420, 2004
- [8] Pedro J. Moreno, Purdy P. Ho "A New SVM Approach to Speaker Identification and Verification Using Probabilistic Distance Kernels" Cambridge Research Laboratory HP Laboratories Cambridge HPL-2004-7 January 9th, 2004
- [9] Vincent Wan and Steve Renals "Speaker verification using sequence discriminate support vector machines" iee transactions on speech and audio processing 2002.
- [10] Shantanu Godbole and Sunita Sarawagi "Discriminative Methods for Multi-labeled Classification" PAKDD 2004, LNAI 3056, pp. 22–30, 2004. Springer-Verlag Berlin Heidelberg 2004
- [11] Reda Jourani, K Daoudi, R_egine Andr_e-Obrecht, Driss Aboutajdine "Combination of SVM and Large Margin GMM modeling for speaker identification" Eusipco, Sep 2013, Marrakech Morocco

Author Profile

Piyush Mishra received a BE in Electronics and telecommunication from Government Engineering college Raipur Chhattisgarh in 2009. Presently he is pursuing M.Tech from SSCET Bhilai Chhattisgarh. His interest includes speech processing machine learning and speech modeling techniques.