Performance of svm, k-nn and nbc classifiers for text-independent speaker identification with and without modelling through merging models

Yussouf Nahayo[*1], Seçkin Arı[2]

ABSTRACT

This paper proposes some methods of robust text-independent speaker identification based on Gaussian Mixture Model (GMM). We implemented a combination of GMM model with a set of classifiers such as Support Vector Machine (SVM), K-Nearest Neighbour (K-NN), and Naive Bayes Classifier (NBC). In order to improve the identification rate, we developed a combination of hybrid systems by using validation technique. The experiments were performed on the dialect DR1 of the TIMIT corpus. The results have showed a better performance for the developed technique compared to the individual techniques.

Keywords: GMM, SVM, KNN, NB, TIMIT

Birleşik modellemeli ve modellemesiz metin-bağımsız konuşmacı tanıma için SVM, K-NN ve NBC sınıflandırıcıların başarımı

ÖZ

Bu çalışma Gaussian Mixture Model tabanlı metin-bağımsız konuşmacı tanıma yöntemleri sunar. GMM model ile Support Vector Machine, K-nearest Neighbour ve Naive Bayes sınıflandırıcı gibi sınıflandırıcıların kombinasyonu gerçekleştirilmiştir. Tanıma oranını iyileştirmek için, doğrulama yöntemi kullanarak hibrid sistemlerin kombinasyonunu geliştirdik. Deneyler TIMIT corpus' un DR1 lehçesi üzerine yapılmıştır. Sonuçlar ayrı ayrı yöntemlerle karılaştırıldığında geliştirilen yöntemle daha iyi başarım göstermiştir.

Anahtar Kelimeler: GMM, Combination, SVM, KNN, NB, TIMIT

* Corresponding Author
[1] Sakarya University, Computer and Information Science, Computer Engineering, Sakarya - nahayoyoussouf@gmail.com
[2] Sakarya University, Computer and Information Science, Computer Engineering, Sakarya - ari@sakarya.edu.tr

Y. Nahayo, S. Arı

Performance of svm, k-nn and nbc classifiers for text independent speaker identification with and without modelling through merging models

1. INTRODUCTION

The human voice is considered as a viable biometric identifier just like a fingerprint or iris. Several efforts are done to increase the performance of biometric person authentification through speech. Speaker identification is one of the most important applications of speaker recognition systems. It is the process of recognizing a speaker among a finite set of speakers by comparing its vocal expression with known references.

Figure 1 represents the basic elements of automatic speaker recognition. A set of discriminative classifiers has found a great attention. In this research, we tested the performance of SVM, K-NN, and Naive Bayes classifier (NBC). The main reason of choosing these classifiers is justified due to the fact that discriminative approaches have been able to dominate the state of art of speaker recognition systems. Thus, the selected classifiers are the most used in automatic speaker recognition and give promising results [1]. GMM is increasingly used to model the feature vectors; also there has been a great interest in combining classifiers in order to improve classification accuracy [2]. The aim of this work consists of comparing the performance of each concerned classifiers, hybrid systems and different strategies of combining the hybrid systems. The remaining of this paper is organized as follows: Section 2 presents the general formulation of the mixture of Gaussians for background modelling. Section 3 describes the different classifiers as well as the strategies of combining them. Section 4 shows a summary of the obtained results.
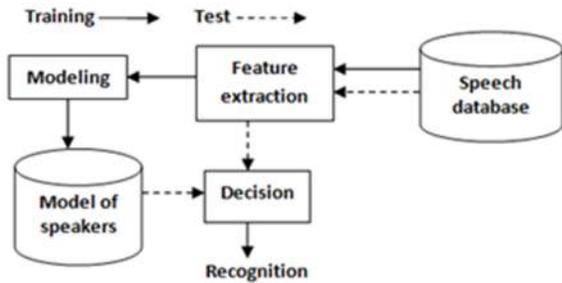

Figure 1. Automatic speaker recognition system

2. GAUSSIAN MIXTURE MODEL BASED SPEAKER IDENTIFICATION SYSTEM

In speaker recognition, there are two types of modelling: The deterministic methods and statistical methods. GMM is among the most statistically mature methods which have become the dominant approach in text-independent speaker identification for its robustness and scalability.
In speaker identification system, usually a Gaussian Universal Background model GMM-UBM approach was

proposed [3] [4] [9]. The UBM is trained using the background databases that are selected to reflect the alternative imposter speeches. The EM algorithm is used for the UBM training [13]. The GMM probability density can be described as follows:

$$p(x) = \sum_{i=1}^{M} w_i f(x \mid m_i, \Sigma_i) \tag{1}$$

where x is a D-dimensional random vector, w_i the weight of the ith Gaussian component, the covariance matrix, m_i the mean vector and $(i = 1,....,M)$. $f(.)$ denotes Gaussian density function i.e.

$$f(x \mid m_i, \Sigma_i) = \frac{(2\pi)^{-D/2}}{|\Sigma_i|^{1/2}} \exp\left\{-\frac{1}{2}(x-m_i)^T \sum_i^{-1}(x-m_i)\right\} \tag{2}$$

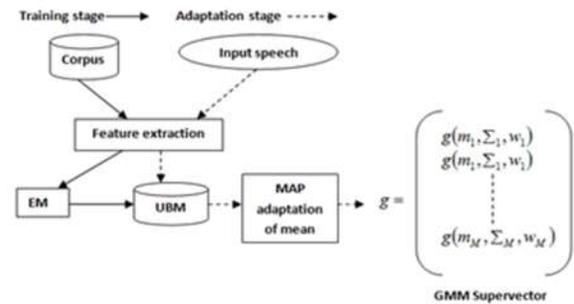The speaker GMM, ϕ can be obtained by MAP adaptation, and it has the same form as follows:

Figure 1. The process of supervectors generations

$g(mi, \Sigma i, wi)$ is the function that represents the normalized mean aligned by covariance and weight. The UBM can be expressed by:

$$y = \left\{w_i^{(y)}, m_i^{(y)}, \Sigma_i^{(y)} \mid i = 1,2,3,......,M\right\} \tag{3}$$

The speaker GMM, ϕ can be obtained by MAP adaptation, and it has the same form as follows:

$$\phi = \left\{w_i^{(\phi)}, m_i^{(\phi)}, \Sigma_i^{(\phi)} \mid i = 1,2,3,......,M\right\} \tag{4}$$

The process of generating the GMM-supervector can be summarized in Figure 2. The GMM-supervector is formed by concatenating the normalized means of the Gaussian components [3]. GMM supervectors will be used in our different hybrid classification system as input vectors for classifiers.

Performance of svm, k-nn and nbc classifiers for text-independent speaker identification with and without modelling through merging models

Y. Nahayo, S. Arı

3. MACHINE LEARNING TECHNIQUES

3.1. Support vector machines

Support vector machines is a supervised technique for solving problems of discrimination, classification, regression, inspired by statistical theory of learning introduced by Vapnik(1995) [3] [5] [8].

SVM is essentially binary nonlinear classifier used to process data with high dimension. Since its introduction in pattern recognition, several studies have emonstrated the effectiveness of this classifier. It can be used for several tasks such as face detection in images, speaker recognition. We briefly present the principle of SVM in two different cases: SVM in linearly separable data case : vectors machine construct a hyper plane that has the largest distance to the nearest training data points of any class and that separates positive examples from negative examples.

SVM in nonlinearly separable data case : The idea is divided into two stages: Transformation of the nonlinear space in new linear space by kernel function and application of a linear SVM classifier [12].

3.2. K-nearest neighbour

K-nearest neighbour (K-NN) classifier is a supervised classification method. It has been used for similarity measure between extracted features and a set of reference features by using Euclidean distance. Given a new instance y, a K-NN classifier finds the K nearest neighbours to the unlabeled data, by computing the distance between the feature vector of the new instance and all feature vectors in the training set. The class for y is estimated as the class which is most represented among the nearest k vectors. Mathematically, this can be described to compute the a posteriori class probability $P(c \mid y)$ as:

$$P(c \mid y) = (k^i / k) * P(c) \qquad (5)$$

where k^i represents the number of vectors belonging to class c within the subset of k vectors.

3.3. Naives Bayes Classifier

Naive Bayes classifier is a supervised classification method which is probabilistic, simple and based on the application of Bayes' theorem: descriptors are pair wise independently-owned, conditional on the values of the variable to predict. This theorem has many applications in information processing including speech processing, image processing, etc...

The probabilistic model for a Bayesian classifier is a conditional model estimated from a set of examples for learning. The classification of a new example is provided by the use of Bayesian decision rule, during the selection of the class with the largest probability. The Naïve Bayesian works as follows:

Given c_n classes and each one has a probability $P(c_n)$ estimated from the training dataset and represents the prior probability of classifying an attribute v_j into c_n.

For n attribute value, v_j, the classification is to find this probability:

$$\frac{P(v_1 \wedge v_2 \dots\dots v_j \mid c_n) P(c_n)}{P(v_1 \wedge v_2 \dots\dots v_j)} \qquad (6)$$

3.4. Combining the methods

In order to get higher prediction accuracy, the idea of combining classifiers has been considered to develop powerful systems in many fields, [6] [10] [11].

3.4.1. Combination of GMM and classifiers (hybrid systems)

The main aim of hybrid systems is to increase the identification rate and reducing the computation time of the recognition system. This is due to the GMM functionality of reducing the classifiers input matrix by transforming the input of thousand frames into input of supervectors as shown in Figure 2. Figure 3 represents the architecture of hybrid systems.
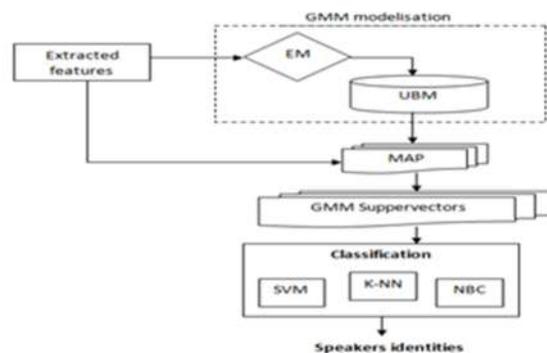

Figure 3. Hybrid systems architecture.

Y. Nahayo, S. Arı

Performance of svm, k-nn and nbc classifiers for text independent speaker identification with and without modelling through merging models

3.4.2. Combination of hybrids classifiers

Recently, research in speaker identification has been moving towards the integration of this strategy. Different methods of combination techniques have been proposed, in our work we opted for the parallel combination of the different machine learning outputs using GMM supervectors as input features [7].

The idea of this combination is shown in Figure 4, This system consists of two main steps: The first step is for the classification where each classifier (SVM, K-NN, NB) operates independently of other classification systems. The decisions of these classifiers are then combined through majority voting mode. For this combination mode, the output of each method is considered as a vote for a class. Number of votes for each class is counted. The class with maximum votes will be retained. [11]
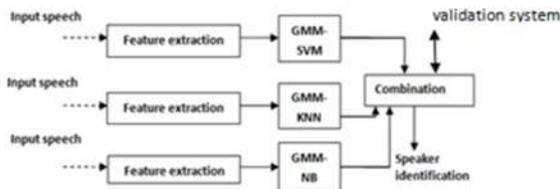

Figure 4. Combination architecture of the hybrid systems

4. EXPERIMENTS

4.1 Utilized Methods

In this study, in the first step we tested the performance of classifiers (SVM, K-NN, NBC) individually. In the second step, we tested their performance by merging them with GMM. Finally, we tested the performance of different combination strategies of these classifiers which is the main contribution of this work. We structured the classifiers combination strategies into different groups as follows:

- GMM-NB+ GMM-K-NN

- GMM-SVM+GMM-K-NN

- GMM-SVM+GMM-NB

- GMM-K-NN+GMM-NB+GMM-SVM

4.2. Corpus

To evaluate the different systems proposed, we used the dialect DR1 (New England dialect) of the TIMIT corpus (18 females and 31 males). Each speaker pronounced 10 sentences, the 8 first sentences are used in the training phase and the last 2 sentences are used in the test phase.

4.3. Experimental conditions

Table 1. shows the different experimental conditions used in our tests.

Table 1. Experimental conditions

Feature extraction	Coefficients : 12 MFCC Sampling frequency: 16 KHz Window length: 16ms sampling interval : 8ms Windowing: Hamming Number of Filter: 24
Modelling	Gaussian number:128 EM iteration :1000
SVM	Kernel :linear Number of iterations for k_Means:100
K-NN	Euclidean distance Number of neighbours nearest=10

4.4. Results

Table 2 bellow shows the classification rates obtained for different classifiers without using GMM modelling.

Table 2. Identification rate of single classifiers without modelling

Without modelling GMM			
Classifiers	**SVM**	**K-NN**	**NB**
Identification rate (%)	9	37	15

From this table, the identification rate is varied between 9% and 37%. The SVM classifier presents the lowest rate. On the other side, K-NN classifier presents the highest rate.

The identification rates of different classifiers are generally low, particularly for SVM classifier, and NB classifier, while the identification rate for K-NN classifier is average by comparing it to the other classifiers, this weakness is explained by the size of the input matrix which is formed by thousands of signal frames.

The K-NN classifier is able to resist moderately to the size of this matrix by giving an average rate of 37%, this is due to its simplicity and strength of classification method.

In the following experiment; after emerging classifiers with GMM, we evaluated the performance of the hybrid systems. The results obtained are shown in Table 3.

Performance of svm, k-nn and nbc classifiers for text-independent speaker identification with and without modelling through merging models

Y. Nahayo, S. Arı

Table 3. Identification rate of the classifiers with GMM modeling : Hybrid system

Without modelling GMM			
Classifiers	SVM	K-NN	NB
Identification rate (%)	96	87	92

By comparing the performance of single classifiers and hybrid classifiers in term of identification rate. The experiments demonstrate that hybrid classifiers give good results which are varied between 87% and 96%, it's interesting to note that even the use of hybrid classifiers yields better results in overall than by using single classifiers, for example single SVM classifier give 9% of identification rate while by merging it with GMM modelling, the identification rate increases towards 96%. This is explained by the ability and the performance of the global modelization approach GMM and the importance of using supervectors.

K-NN classifier algorithm is one of the simplest algorithms of automatic supervised study, but is very important to determine the number of k nearest neighbours which provide a good performance of identification rate. after a series of tests, the number of nearest neighbours was fixed at 10 as the value that gives optimal performance as is shown on the Figure 5.

The performance results of combining these hybrid classifiers in different strategies is shown in Table 4.

The same number of coefficients was used for all classifiers, 12 MFCC coefficients. For the number of Gaussians, we applied 128 for all tests and k nearest neighbours number of K-NN approach we use 10 for all strategies of combination where K-NN is concerned.
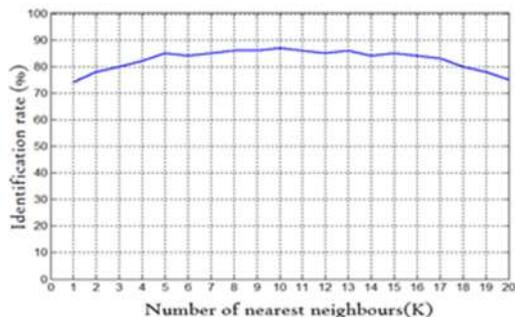

Figure 5. The variation of identification rate vs the number of nearest neighbours.

Table 4. shows that the identification rate is varied between 97% and 100%. These strategies of combining hybrid classifiers give better results than the individual classifiers that can even achieve an interested

identification rate of 100% for strategy 4 after running the system several time. We can conclude that combining hybrid classifiers is an effective solution to the problem of speaker identification.

Table 4. Results of the combined classifiers

Strategies	Identification rate (%)
Strategy1: GMM-NB+GMM-K-NN	97
Strategy2: GMM-SVM+GMM-K-NN	96
Strategy3: GMM-SVM+GMM-NB	98
Strategy4: GMM-K-NN+GMM-NB+GMM-SVM	100

5. CONCLUSION

In this paper, we implemented different discriminative approaches systems(SVM, K-NN, NBC) with and without GMM modelling and we proposed a combination of hybrid systems in order to enhance the performance of identification for text independent speaker's identification system.

Experimental results have shown that hybridization of classifiers (SVM,K-NN,NB) with GMM and combination methods of these classifiers bring a significant performance over the single classifier. Indeed, the different combination strategies present an interested improvement of the identification rate, which can even reach 100% for the strategy .

As perspectives, we will try to integrate one or several modalities (such as lips movement, face picture, etc) to the speech and merge them to characteristic parameters in order to test the effectiveness of our combination hybrid systems in front of a large dataset.

REFERENCES

[1] D. A. N. R.Amami, "An Empirical Comparison of SVM and Some Supervised Learning Algorithms for Vowel recognition", International Journal of Intelligent Information Processing IJIIP, 2012.

[2] B.S. Atal, "Automatic Recognition of Speaker from Their Voices", Proceedings of the IEEE, Vol. 64, No. 4, pp 460-475, 1976

[3] W. M. Campbell, D. E. Sturim, D. A. Reynolds, and A. Solomon off, "SVM based speaker verification using a GMM supervector kernel and

Y. Nahayo, S. Arı

Performance of svm, k-nn and nbc classifiers for text independent speaker identification with and without modelling through merging models

NAP variability compensation", Proc. Int. Conf. Acoustics, Speech, and Signal Processing, 2006.

[4]   D. Reynolds and R. Rose, "Robust text-independent   speaker identification using Gaussian mixture speaker models, " IEEE Trans. Speech Audio Proc., vol. 3, no. 1, pp. 72–83, 1995.

[5]   D. Ben Ayed Mezghani, S. Zribi Boujelbene et N. Ellouze, "Evaluation of SVM kernels function and conventional machine learning algorithms for Speaker Identification Task," International Journal of Hybrid Information Technology (IJHIT), vol. 3, pp. 2 3-34, 2010.

[6]   S. Zribi Boujelbene, D. Ben Ayed Mezghan et N. Ellouze, "Application of Combining Classifiers for Text-Independent Speaker Identification," the 16th IEEE International Conference on Electronics, Circuits, and Systems ICECS, Hammamet-Tunisie, pp. 723-726, 2009.

[7]   R. Djemili, M. Bedda and H. Bourouba, "A Hybrid GMM/SVM System for Text Independent Speaker Identification," International Journal of Computer and Information Science and Engineering, vol. 1, pp. 1-8, 2007.

[8]   D. Neiberg "Text Independent Speaker Verification Using Adapted Gaussian Mixture Models", Centre for Speech Technology (CTT) Department of Speech, Music and Hearing KTH, Stockholm, Sweden 2001-12-11.

[9]   S. Zribi Boujelbene, D. Ben Ayed Mezghan et N. Ellouze, "Support Vector Machines approaches and its application to speaker identification," 3rd IEEE International Conference on Digital Ecosystems and Technologies DEST, pp. 662-667, 2009.

[10] I. Ayed, "Stratégies de fusion de paramètres pour une tâche d'identification du locuteur en mode indépendant du texte : Application sur le corpus NTIMIT.TAIMA", Hammamet-Tunisie 2011.

[11] L. Lam et C.Y. Suen. "Application of Majority Voting to Pattern Recognition: An Analysis of Its Behavior and Performance", IEEE Transactions on Systems, Man Cybernetics, pp. 553-568, 1997.

[12] K. S. Durgesh, and B. Lekha, "Data classification using support vector machine." Journal of Theoretical and Applied Information Technology, 12(1), 1-7, 2010

[13] H. Y. Chang, A. L. Kong, and L. Haizhou, "An SVM Kernel With GMM-Supervector Based on the Bhattacharyya Distance for Speaker Recognition", v.6, pp. 1300-1312, 2010