IFAC Publications

SPEAKER VERIFICATION BASED ON FUZZY INFERENCE SYSTEM WITH MOVING PARAMETRIC CONSEQUENTS

Adam Dustor *

* *Institute of Electronics, Silesian University of Technology, Akademicka 16, 44-100 Gliwice, Poland, adam.dustor@polsl.pl*

Abstract: This article presents an application of fuzzy nonlinear classifier to voice biometrics namely to speaker verification task. This classifier is closely related to a modification of a classical Takagi-Sugeno-Kang inference system and is based on a fuzzy moving consequents in If-Then rules. Good generalization properties of this classifier enable to achieve low error rates even for short training speaker utterances. Achieved verification results are compared with the ones obtained for the most popular in speaker recognition area techniques like Gaussian Mixture Models and vector quantization. All research is based on Polish speech corpus ROBOT designed for testing speech algorithms.

Keywords: signal processing, speech analysis, speaker verification, speaker identification, pattern recognition,

1. INTRODUCTION

Speaker verification belongs to one of many biometric techniques. Its aim is to decide whether the speaker is whom he claims to be. Main applications of speaker verification are secure access to services via telephone, home banking and verification in WWW applications.

The task of speaker verification consists of several steps. At first speech, after acquisition, is converted into sequence of multidimensional vectors. This step is known as a feature extraction. Next, pattern matching is done during which similarity between speaker model and the sequence of features extracted from recognized utterance is computed. Since verification is a binary process, it is necessary to make an accept/reject decision on the basis of computed similarity score. The last but not the least step is an enrolment responsible for generating speaker reference models known also as a training of the system. The performance of verification process is presented by means of DET curves (Martin *et al.*, 1997).

The paper is organized in the following way. Section 2 presents briefly classical methods used to speaker model construction. Section 3 discusses fuzzy nonlinear classifier based on fuzzy inference system with moving parametric consequents in If-Then rules and the classical Ho-Kashyap procedure. Section 4 describes corpus ROBOT applied for research and finally in section 5 description of automatic speaker verification system in Matlab environment and obtained results of verification accuracy are presented. Summary of achieved results is in section 6.

2. SPEAKER MODELS

As speaker verification is based on similarity calculation between test utterance and reference model, it is obvious that the problem of speaker model construction is crucial. There are two broad families of speaker models namely generative and discriminative.

Generative models are the probability density estimators that attempt to capture all of the underlying fluctuations and variations of the speaker's voice. These models include Gaussian mixture models GMM (Reynolds, 1995; Dustor, 2003*a*), Gaussian classifier GC and the broad family of the nearest neighbor classifiers based on vector quantization VQ techniques (Dustor, 2003*b*; Dustor, 2003*c*).

Discriminative models are optimized to minimize the error on a set of training samples. From this category an application in speaker recognition found support vector machines SVM (Vapnik, 1995). Since discriminative approach should theoretically yield better performance than generative (Vapnik, 1995) it is very interesting to test in speaker recognition the performance of other discriminative classifiers. Especially interesting is an application of classifiers leading to the lowest errors on a standard data sets used in pattern recognition like Kernel Ho-Kashyap classifier KHK (Łęski, 2004; Dustor, 2004*b*) and presented in this paper fuzzy nonlinear classifier based on fuzzy inference system with moving parametric consequents in If-Then rules. This classifier denoted shortly FHK (*Fuzzy Ho-Kashyap*) is a modification of a classifier presented in (Łęski, 2003; Dustor, 2004*a*).

3. FUZZY NONLINEAR CLASSIFIER

The classifier is designed on the basis of the training set, $Tr = \{(\mathbf{x}_1, y_1), (\mathbf{x}_2, y_2), \cdots, (\mathbf{x}_N, y_N)\}$ where $\mathbf{x}_i \in \mathcal{R}^t$ is a feature vector extracted from a frame of speech, N is the number of these vectors and $y_i \in \{-1, +1\}$ indicates the assignment to one of two classes ω_1 or ω_2. After defining the augmented vector $\mathbf{x}_i^{'} = [\mathbf{x}_i^T, 1]^T$ the decision function of the classifier can be defined as

$$g(\mathbf{x}_i) = \mathbf{w}^T \mathbf{x}_i^{'} \begin{cases} \geq 0, & \mathbf{x}_i \in \omega_1, \\ < 0, & \mathbf{x}_i \in \omega_2, \end{cases} \tag{1}$$

where $\mathbf{w} = [\widetilde{\mathbf{w}}^T, w_0]^T \in \mathcal{R}^{t+1}$ is a weight vector which must be found during training of the classifier. After multiplying by -1 all patterns from ω_2 class the equation (1) can be rewritten in the form $y_i \mathbf{w}^T \mathbf{x}_i^{'} > 0$ for $i = 1, 2, \cdots, N$. Let $\mathbf{X}$ be the $N \times (t+1)$ matrix

$$\mathbf{X} = \begin{bmatrix} y_1 \mathbf{x}_1^{'T} \\ y_2 \mathbf{x}_2^{'T} \\ \vdots \\ y_N \mathbf{x}_N^{'T} \end{bmatrix}, \tag{2}$$

then (1) can be written in the matrix form $\mathbf{Xw} > 0$. To obtain solution $\mathbf{w}$ this inequality is replaced by $\mathbf{Xw} = \mathbf{b}$ where $\mathbf{b} > 0$ is an arbitrary vector called a classifier margin. If data are linearly separable then all components of error vector $\mathbf{e} = \mathbf{Xw} - \mathbf{b}$ are greater than zero and by increasing the respective component of $\mathbf{b}$ (b_p) the value of e_p can be set to zero. If $e_p < 0$ then the p-th pattern $\mathbf{x}_p$ is wrongly classified and it is impossible to retain the condition $b_p > 0$ while decreasing b_p. As a result the misclassification error can be written in the form

$$J(\mathbf{w}, \mathbf{b}) = \sum_{i=1}^{N} \mathcal{H}(-e_i), \tag{3}$$

where $\mathcal{H}(\bullet)$ is the unit step pseudo-function, $\mathcal{H}(e_i) = 1$ for $e_i > 0$ and $\mathcal{H}(e_i) = 0$ otherwise. Obtaining solution $\mathbf{w}$ requires minimization the criterion (3). Unfortunately due to its non-convexity, criterion (3) must be approximated by

$$J(\mathbf{w}, \mathbf{b}) = \sum_{i=1}^{N} |e_i|, \tag{4}$$

or

$$J(\mathbf{w}, \mathbf{b}) = \sum_{i=1}^{N} (e_i)^2. \tag{5}$$

Better approximation of (3) and more robust to outliers is the criterion (4).

Vectors $\mathbf{w}$ and $\mathbf{b}$ are found by minimization the function (Łęski, 2003)

$$J(\mathbf{w}, \mathbf{b}) = (\mathbf{Xw} - \mathbf{b})^T \mathbf{D} (\mathbf{Xw} - \mathbf{b}) + \tau \widetilde{\mathbf{w}}^T \widetilde{\mathbf{w}}, \tag{6}$$

where matrix $\mathbf{D} = \mathrm{diag}(d_1, d_2, \cdots, d_N)$ and d_i is the weight corresponding to the i-th pattern, which can be interpreted as a reliability attached to this pattern. The second term of (6) is responsible for the minimization of the complexity of the classifier. The regularization constant $\tau > 0$ controls the trade-off between the classifier complexity and the amount up to which the errors are tolerated. The optimum value of τ is found by cross-validation on the test set.

Differentiation of (6) with respect to $\mathbf{w}$ and $\mathbf{b}$ and setting the results to zero yields the conditions (Łęski, 2003)

$$\begin{cases} \mathbf{w} = (\mathbf{X}^T \mathbf{D} \mathbf{X} + \tau \widetilde{\mathbf{I}})^{-1} \mathbf{X}^T \mathbf{D} \mathbf{b}, \\ \mathbf{e} = \mathbf{Xw} - \mathbf{b} = 0, \end{cases} \tag{7}$$

where $\widetilde{\mathbf{I}}$ is the identity matrix with the last element on the main diagonal set to zero. Vector $\mathbf{w}$ depends on margin vector $\mathbf{b}$. If pattern lies on the right side of the separating hyperplane then corresponding margin can be increased to obtain