International Symposium on Chinese Spoken Language Processing (ISCSLP 2004)
Hong Kong
December 15-18, 2004

A METHOD OF ESTIMATING THE EQUAL ERROR RATE FOR AUTOMATIC SPEAKER VERIFICATION

Jyh-Min CHENG and Hsiao-Chuan WANG

Department of Electrical Engineering,
National Tsing Hua University, Hsinchu 300, Taiwan, R.O.C.
d927908@oz.nthu.edu.tw, hcwang@ee.nthu.edu.tw

ABSTRACT

In an automatic speaker verification (ASV) system, the equal error rate (EER) is a measure to evaluate the system performance. Usually it needs a large number of testing samples to calculate the EER. In order to estimate the EER without running the experiments using testing samples, a method of model-based EER estimation which computes likelihood scores directly from client speaker models and imposter models is proposed. However, the distribution of the computed likelihood scores is significantly biased against the distribution of likelihood scores obtained from testing samples. Here we propose a novel idea to manipulate the speaker models of the client speakers and the imposters so that the distribution of the computed likelihood scores is closer to the distribution of likelihood scores obtained from testing samples. Then a more reliable EER can be calculated by the speaker models. The experimental results show that the proposed method can properly estimate the EER.

1. INTRODUCTION

The automatic speaker verification (ASV) refers to the task of verifying the speaker's identity from their voices. In the conventional ASV system, the decision rule of acceptance or rejection is based on the score of a testing utterance and a predefined threshold [1]. There are two phases for an ASV system to accomplish this task. In the training phase, it learns each client's voice features from the training utterances to generate a statistical model. Also, a single speaker independent model called the world model or the universal background model (UBM) [2] is generated. In the testing phase, the ASV system analyzes an incoming utterance, and then uses the claimed speaker model and the UBM to compute a log-likelihood ratio (LLR) score. Then the score is compared with a preset threshold to determine whether the speaker is accepted or rejected. The Gaussian mixture model (GMM) method has been successfully used for this purpose [1][3].

The speech data of a specific speaker is trained with EM method [4] to derive the corresponding GMM. The LLR score is based on computing the log likelihood ratio of the client speaker model and the UBM. The score distributions are then used to compute the false rejection rate (FRR) and the false acceptance rate (FAR). The equal error rate (EER), where FRR equals to FAR, is a common measure of the performance of the ASV system. Usually, we need a large number of testing samples to evaluate the performance of the ASV system, i.e., to compute the EER.

A model-based framework of classification error rate estimation has been proposed for automatic speech recognition [5]. It aims at predicting the run-time performance of hidden Markov model (HMM) based recognition system for a given task vocabulary and grammar without the need of running recognition experiments using a separate set of testing samples. This motivates the idea of evaluating the performance of an ASV system without running the speaker verification experiments.

In this paper, we derive a method for estimating the EER of an ASV system directly by the parameters of the corresponding GMMs. It computes LLR scores of the true speakers and the imposters, and then calculates the FRR and the FAR. If we used an approximation in [5] to compute mixture densities in the log-of-sum of the integration of LLR score, the distribution of the LLR scores using the model-based method is significantly biased against the distribution of scores obtained by running the verification experiments with a large number of testing samples. This observation motivates the development of an alternative method to manipulate the client and imposter models, so that the distribution of the computed likelihood scores is closer to the distribution of likelihood scores obtained from testing samples. Then, we derive an estimation of the EER which is close to the EER obtained from running speaker verification experiment with a large number of testing samples.

In the following sections, we will first introduce the computation of the LLR scores directly by the model parameters. Secondly, we show the observations of the distribution of the model-based LLR scores based on NIST 19999 Speaker Evaluation data. Thirdly, we show how we manipulate the speaker models of the client speakers and the imposters. Finally, we demonstrate an experiment showing that we can get a close estimation of the EER of the ASV system.

2. MODEL-BASED LLR SCORES

This section introduces the algorithm for estimating the LLR score in terms of the model parameters without using a large number of testing samples.

2.1. Definitions

A log likelihood ratio (LLR) score is defined as the following equation related to the test speaker model p, the client speaker model q, and the UBM r.,

$$\Lambda(p;q,r)=\int_{\mathbf{x}} p(\mathbf{x})\log\frac{q(\mathbf{x})}{r(\mathbf{x})}d\mathbf{x} \tag{1}$$

where $\mathbf{x}$ is the incoming feature vector. Equation (1) can also be expressed as,

$$\begin{aligned}&\Lambda(p;q,r) \\ &=\int_{\mathbf{x}} p(\mathbf{x})\log q(\mathbf{x})d\mathbf{x}-\int_{\mathbf{x}} p(\mathbf{x})\log r(\mathbf{x})d\mathbf{x} \\ &=L(p;q)-L(p;r)\end{aligned} \tag{2}$$

where $L(\bullet)$ is the log likelihood score.

2.2. Derivation of the LLR scores

The log likelihood score $L(\bullet)$ in the mixture model is computed by,

$$\begin{aligned}L(p;q)&=\sum_{i=1}^{M} w_{pi}\int_{\mathbf{x}} N(\mathbf{x};\boldsymbol{\mu}_{pi},\mathbf{S}_{pi}) \\ &\cdot\log\left(\sum_{j=1}^{M} w_{qj}N(\mathbf{x};\boldsymbol{\mu}_{qj},\mathbf{S}_{qj})\right)d\mathbf{x}\end{aligned} \tag{3}$$

where $N(\bullet)$ is the Gaussian density and M is the corresponding number of mixtures. w, $\boldsymbol{\mu}$, $\mathbf{S}$ are the mixture weight, mean, and covariance, respectively.

In order to solve the difficulty in computing the log-of-sum, we make an approximation,

$$\begin{aligned}L(p;q)&=\sum_{i=1}^{M} w_{pi}\log\left(\sum_{j=1}^{M} w_{qj}\right. \\ &\left.\cdot\exp\left(\int_{\mathbf{x}} N(\mathbf{x};\boldsymbol{\mu}_{pi},\mathbf{S}_{pi})\cdot\log(N(\mathbf{x};\boldsymbol{\mu}_{qj},\mathbf{S}_{qj}))d\mathbf{x}\right)\right)\end{aligned} \tag{4}$$

The last term in the exponential of equation (4) can be then computed as the expected value of $\log(N(\mathbf{x};\boldsymbol{\mu}_{qj},\mathbf{S}_{qj}))$ given the distribution of $N(\mathbf{x};\boldsymbol{\mu}_{pi},\mathbf{S}_{pi})$,

$$\begin{aligned}&E_{pi}\left[\log(N(\mathbf{x};\boldsymbol{\mu}_{qj},\mathbf{S}_{qj}))\right] \\ &=-\frac{D}{2}\log(2p)-\frac{1}{2}\log\left(\left|\mathbf{S}_{qj}\right|\right) \\ &-\frac{1}{2}E_{pi}\left[(\mathbf{x}-\boldsymbol{\mu}_{qj})^{T}\mathbf{S}_{qj}^{-1}(\mathbf{x}-\boldsymbol{\mu}_{qj})\right]\end{aligned} \tag{5}$$

where D is the vector dimension, and the last term in equation (5) can be simplified by,

$$\begin{aligned}&E_{pi}\left[(\mathbf{x}-\boldsymbol{\mu}_{qj})^{T}\mathbf{S}_{qj}^{-1}(\mathbf{x}-\boldsymbol{\mu}_{qj})\right] \\ &=tr\left(\mathbf{S}_{pi}\mathbf{S}_{qj}^{-1}\right)+(\boldsymbol{\mu}_{pi}-\boldsymbol{\mu}_{qj})^{T}\mathbf{S}_{qj}^{-1}(\boldsymbol{\mu}_{pi}-\boldsymbol{\mu}_{qj})\end{aligned} \tag{6}$$

where $tr(\bullet)$ is the trace of matrix $(\bullet)$. Then, the log likelihood score can be computed by the following equation,

$$\begin{aligned}L(p;q)&=\sum_{i=1}^{M} w_{pi}\log\left(\sum_{j=1}^{M} w_{qj}\frac{1}{(2p)^{D/2}\left|\mathbf{S}_{qj}\right|^{1/2}}\right. \\ &\left.\cdot\exp\left(-\frac{1}{2}\left(tr\left(\mathbf{S}_{pi}\mathbf{S}_{qj}^{-1}\right)+(\boldsymbol{\mu}_{pi}-\boldsymbol{\mu}_{qj})^{T}\mathbf{S}_{qj}^{-1}(\boldsymbol{\mu}_{pi}-\boldsymbol{\mu}_{qj})\right)\right)\right)\end{aligned} \tag{7}$$

By equation (2) and (7), we could compute the LLR score directly by the model parameters without running the verification experiments.

3. MODEL-BASED EER ESTIMATION

The EER estimation of an ASV system is based on the computation of LLR scores. Usually, a large number of testing samples is necessary for evaluating FRR and FAR, so that the EER can be determined. Here, we try to evaluate EER by using model parameters.

3.1. FRR vs. FAR

In an ASV system, there are two types of error, i.e., false rejection (FR) and false acceptance (FA). In equation (1), q and r are the claimed speaker model and the UBM, respectively. For the FR, the test speaker model p is very similar to the claimed speaker model q,

$$\Lambda(p;q,r)=\int_{\mathbf{x}} q(\mathbf{x})\log\frac{q(\mathbf{x})}{r(\mathbf{x})}d\mathbf{x} \tag{8}$$

For the FA, the test speaker model p is the imposter mixture densities, we approximate it as the UBM r,

$$\Lambda(p;q,r)=\int_{\mathbf{x}} r(\mathbf{x})\log\frac{q(\mathbf{x})}{r(\mathbf{x})}d\mathbf{x} \tag{9}$$

We can observe that the resulted value of (8) is positive and that of (9) is negative. Figure 1 shows the distributions of the LLR scores either derived from running verification experiments using a large number of training samples or derived from the trained models.

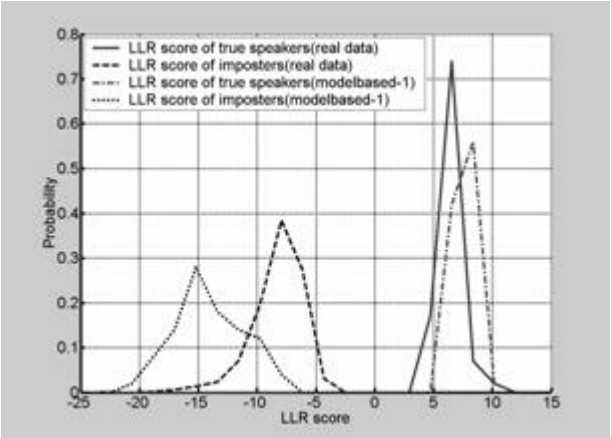

Figure 1: Distribution of data-based LLR scores

From figure 1, we found that the distributions of the LLR scores derived from the model parameters are different from those obtained from running the experiment with training data. Let the test speaker model p be replaced by a new defined model $\bar{q}$. Equation (9) becomes,

$$\Lambda(p;q,r)=\int_{\mathbf{x}}\bar{q}(\mathbf{x})\log\frac{q(\mathbf{x})}{r(\mathbf{x})}d\mathbf{x} \qquad (10)$$

If $\bar{q}$ is derived from a set of the client speakers except the claim speaker model q , the distributions of the LLR scores become Figure 2.

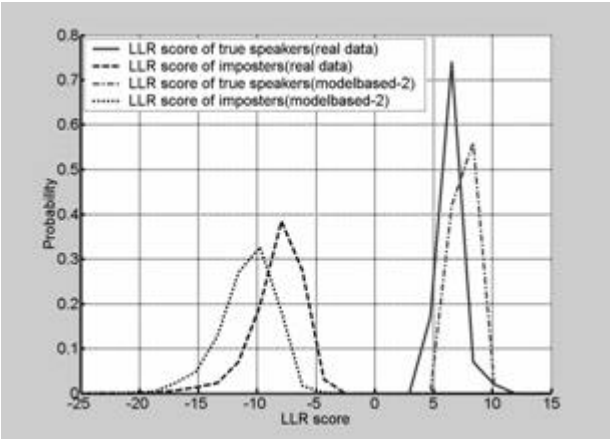

Figure 2: Distributions of model-based LLR scores

Figure 2 shows that two distributions of LLR scores of imposters are closer. This gives us an idea to manipulate the speaker mo dels of the true speakers and the imposters.

3.2. Manipulating the test speaker model

Let's define an LLR score as follows.

$$\Lambda(p;q,r)=\int_{\mathbf{x}}((1-c)\bar{q}(\mathbf{x})+cq(\mathbf{x}))\log\frac{q(\mathbf{x})}{r(\mathbf{x})}d\mathbf{x} \qquad (11)$$

where c is a weighting factor.

Let $\bar{q}$ be a set of N speaker models which are close to the claim speaker model q, i.e., $\bar{q}=\min_{q'\neq q}KLD(q',q)$, where KLD is the Kullback-Leibler distance [5],

$$KLD(q,q')=\Lambda(q;q,q')+\Lambda(q';q',q) \qquad (12)$$

Then we can properly choose the weighting factor c and N client speaker models which are close to the claimed speaker q to calculate the LLR score for claimed speaker q.

4. EXPERIMENTS

4.1. Database

In our experiments, we use the male subset of the evaluation data of the NIST 1999 Speaker Verification task. It provides 230 male speakers. Each speaker has two minutes training data with two different channels recorded in 8 kHz sampling rate and ITUT G.711 μ-law encoded format. In the testing set, there are 1477 utterances for male speakers.

4.2. Experimental arrangement

The speech data is pre-emphasized with a factor of 0.97. The analysis window is of 32 ms length. The frame shift is 16 ms. 15-order Mel-frequency cepstral coefficients (MFCCs) extracted from each frame form a feature vector for generating speaker models. Speaker models are represented in GMM of 1024 mixture components. The UBM is generated using data of 60 male speakers in NIST 1999 Training data.

4.3. Experimental results

The experiments are arranged to investigate the goodness of the model-based EER estimation. This can be accomplished by measuring the difference between the model-based EER estimation and the EER obtained by running the verification experiment using a large number of testing samples, i.e., the whole testing samples. The EER measured by running the verification experiment using whole testing samples is considered as a result of reference experiment. For the performance evaluation, the LLR scores computed by using equations (8) and (9) is denoted as modelbased-1, and the LLR scores computed by using equations (10), (11) and (12) is denoted as modelbased-2. From figure 3 and figure 4, we can observe that the distributions of LLR scores in figure 4 are closer to that of the reference experiment.

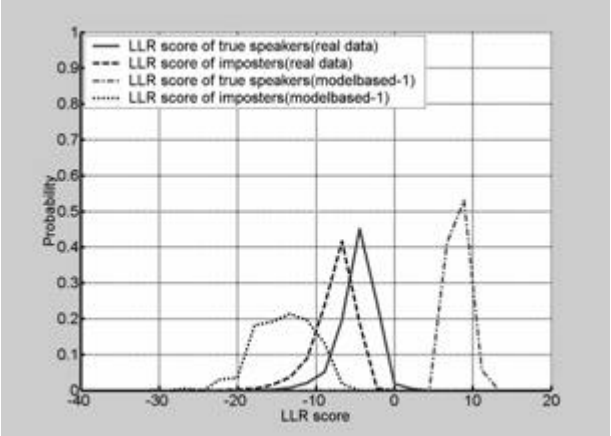

Figure 3: modelbased-1

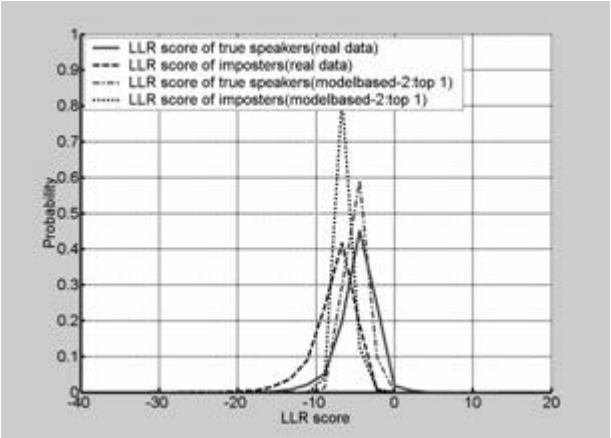

Figure 4: modelbased-2

In order to investigate the effectiveness of the weighting factor in equation (11), an experiment with different weighting factor c is shown in figure 5. We can observe that the weighting factor c increases as N increases. Since we choose more client speaker models into $\bar{q}$, the distribution of the LLR scores becomes closer to that of the imposters.

The experimental results of the comparisons between the reference experiment and the modelbased-2 are shown in the table 1. We can see that the reasonable choice of c is around 0.21

Table 1: EERs of reference experiment and the modelbased-2

	reference	top 1	top 5	top 10
EER(%)	23.76	23.70	23.74	23.78
weight	-	0.21	0.214	0.22

5. CONCLUSIONS

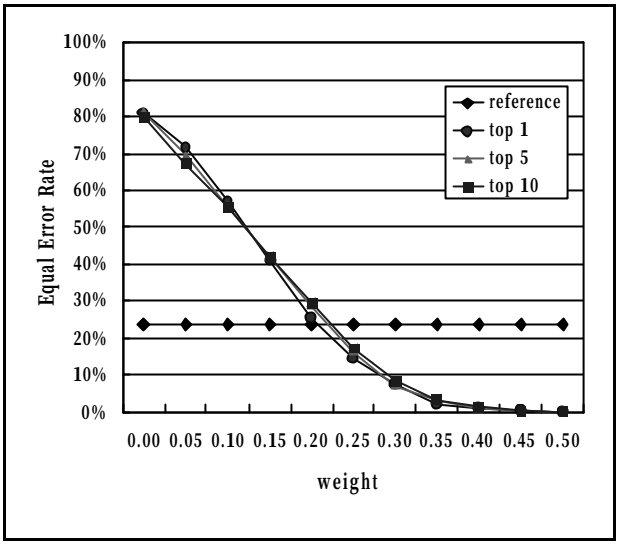

Figure 5: modelbased-2 with different weighting factor c

A preliminary study on the model-based EER estimation for the ASV system is presented. From the experiment, we find some cues to manipulate the speaker models for computing the LLR scores based on speaker models. Experimental results show that we can effectively estimate the performance of the ASV system based on model parameters with a properly chosen factor c. This technique can be used for feedback loop design in an ASV system.

6. ACKNOWLEDGEMENTS

This research was partially sponsored by the National Science Council, Taiwan, under contract number NSC-92-2213-E-007-036

7. REFERENCES

[1] Reynolds D., "Speaker identification and verification using Gaussian mixture models," Speech Communication, Vol. 17, pp. 91-108, 1995.

[2] D. A. Reynolds, "Comparison of background normalization methods for text-independent speaker verification," Proceedings of the EUROSPEECH, Rhodes, Greece, Sept. 19997.

[3] Reynolds D., Rose R.C., "Robust text-independent speaker identification using Gaussian mixture speaker models," IEEE Transactions on Speech and Audio Processing, Vol. 3m pp. 72-83, Jan. 1995.

[4] Guorong Xuan, Wei Zhang, Peiqi Chai, "EM algorithm of Gaussian mixture and hidden Markov model," IEEE International Conference on Image Processing, Vol. 1, pp. 145-148, 2001.

[5] C.-S. Huang, H.-C. Wang and C.-H. Lee, "A Study on Model-Based Error Rate Estimation for Automatic Speech Recognition," IEEE Transaction on Speech and Audio Processing, Vol. 11, No. 6, Nov. 2003.