Nist Speaker Recognition Evaluation 2006

LRDE - EPITA

R. Dehak, C. Deledalle, N. Dehak

reda@lrde.epita.fr

This work was done in collaboration with TSI-ENST and Computer Science Department of Fribourg university.

Three different systems have been elaborated for the 2006 NIST evaluation for the **1conv4w-1conv4w** task. For all systems, the GMM models have been built using BECARS[1] software. The three speaker recognition systems differ on acoustic features, GMM dimension and scoring approach.

The first two systems used the same database to build the UBM. We have used the same subset as Fribourg university. This subset was extracted from NIST-03 and Fisher databases. The third one have been elaborated during the first Biosecure Residential Workshop[2] and use a subset of NIST-2003 and NIST-2004 databases.

1 Primary system (*LRDE-EPITA_1*)

The primary system use an UBM build with a subset extracted from fisher database. A MAP adaptation was used to adapt UBM to target speaker. This task is elaborated using BECARS software.

First, we removed silence segments. Then, a global normalization of the mean and standard deviation is done on speech segments for all features. We used SPRO[3] to evaluate this feature.

[1] http://www.tsi.enst.fr/becars
[2] http://www.biosecure.info
[3] http://irisa.fr/metiss/spro

For each speech segment, we evaluate a score using log likelihood ratio of the 20^{th} top gaussian's of the UBM and target model. Finally a zt-normalization is done on the decision score.

2 Second system (*LRDE-EPITA_2*)

Before GMM learning, this system uses a feature warping to warp the parameters to a gaussian distribution.

The scoring process is based on the discrimination power of support vectors machines (SVM) in combination with gaussian mixture models. We use a kernel which allows computation of similarity between GMM models, calculated using the distance between UBM and target models.

3 Third system (*LRDE-EPITA_3*)

The third system differs from the two first one in preprocessing step, type and dimension of acoustic features and GMM dimension. It uses the *ctm* files provided by NIST to detect speech segment. A CMS (Cepstral Mean Substraction) has been applied on speech segment of each file before the feature extraction step.

As the second system, we have apply SVM in GMM space to evaluate the decision score. Our motivation is, to find discriminative boundary between target GMM super vector and impostors GMMs super vectors. The kernel used in this SVM is based on distance between GMMs. We have also used a model normalization called M-norm [1] to normalize GMMs. More details about this method can be found in [2].

4 Systems design features

All computations were done on a cluster of 17 nodes, each node has a P4 3GHz CPU, 1GByte of memory, and 1GByte of swap.

4.1 Features extraction

1. ***LRDE-EPITA_1*** :

   LFCC based acoustic representation.

33 features : 16 LFCC, delta LFCC (16), delta energy.
mean and standard deviation normalization of features

2. ***LRDE-EPITA_2*** :

   LFCC based acoustic representation.
   33 features : 16 LFCC, delta LFCC (16), delta energy (1).
   Gaussian features warping.

3. ***LRDE-EPITA_3*** :

   MFCC based acoustic representation.
   31 features : 15 MFCC, delta MFCC (15), delta energy (1).
   speech detection using ctm files.

4.2 Universal Background Model

The first two systems use the same database to build the UBM. We have used the same subset as Fribourg university. This subset was extracted from NIST-03 and Fisher databases. The third one used a subset extracted from NIST-03 and NIST-04 databases. The parameters of all UBMs were estimated using the EM algorithm.

1. ***LRDE-EPITA_1***:

GMM dimension	:	512
Male	:	Nist-03 and Fisher
		749 segments (one-side)
		Computation 144h
Female	:	Nist-03 and Fisher
		807 segments (one-side)
		Computation 150h

2. ***LRDE-EPITA_2***:

GMM dimension	:	2048
Male	:	Nist-03 and Fisher
		749 segments (one-side)
		Computation 232h
Female	:	Nist-03 and Fisher
		807 segments (one-side)
		Computation 240h

3. ***LRDE-EPITA_3***:

GMM dimension	:	512
Male	:	Nist-03 and Nist-04
		200 segments (one-side)
		Computation 24h
Female	:	Nist-03 and Fisher
		283 segments (one-side)
		Computation 30h

4.3 Adaptation

We use means adaptation of world model using MAP algorithm. The first system uses the adaptation to evaluate only targets and impostors models. The last two systems calculate the targets, the imposters and the tests models.

Computation :

LRDE-EPITA_1	:	12h
LRDE-EPITA_2	:	65h for targets and impostors models
		202h for test models
LRDE-EPITA_3	:	11h for target and impostors models
		53h for test models

4.4 Imposteur

The first system uses a set of 373 males impostors and 373 females impostors for score normalization extracted from Fisher databases. The second system uses the same set to evaluate the discriminative boundary between target GMM supervector and impostors GMMs super vectors. The last system use a subset of 195 males impostors and 294 females impostors extracted from NIST-04 and NIST-03 databases.

4.5 Scoring

The first system uses log likelihood ratio of the 20^{th} top gaussian's of the UBM and target model to evaluate the score. The last two systems use two different kernels for the learning and test step.

Computation :

LRDE-EPITA_1	:	8h for scoring and 122h for normalization
LRDE-EPITA_2	:	18h for kernels computation and score evaluation
LRDE-EPITA_3	:	12h for kernels computation and score evaluation

References

[1] M. Ben, *Approches Robustes pour la Vérification Automatique du Locuteur par Normalisation et Adaptation Hirarchique*, Ph.D. thesis, University of Renne I, 2004.

[2] Najim Dehak and Gérard Chollet, "Support Vector GMMs for Speaker Verification," in *IEEE odyssey*, San Juan, Puerto Rico, 2006.