Acoustic Vector Resampling for GMMSVM-Based Speaker Verification

Man-Wai MAK and Wei RAO

Dept. of Electronic and Information Engineering, The Hong Kong Polytechnic University

Introduction

This paper proposes a resampling technique to mitigate the data imbalance problem in GMMSVM-based speaker verification. The sequence order of acoustic vectors in an enrollment utterance is first randomized; then the randomized sequence is partitioned into a number of segments. Each of these segments is then used to produce a GMM-supervector. A desirable number of speaker-class supervectors can be produced by repeating this randomization and partitioning process a number of times. Evaluations suggest that the method can reduce the EER of GMM-SVM systems by 10%.

Motivation

A problem in SVM scoring is that the number of target speaker utterances for training the target-speaker's SVM is very limited (typically only one enrollment utterance is available).

Given that the number of background speakers's utterances is typically several hundreds, the limited number of enrollment utterances leads to a severe **data imbalance problem**. An undesirable consequence of data imbalance is that the orientation of the decision boundary is largely dictated by the data in the majority (background speakers) class.

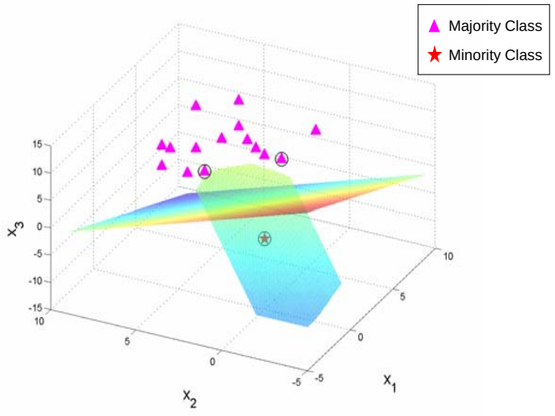

Fig.1: *A 3D two-class problem illustrating the imbalance between the number of minority-class samples (red star) and majority-class samples (pink triangles) in a linear SVM. The decision plane is defined by 3 support vectors (enclosed by black circles). The green region beneath the decision plane represents the region where the minority-class sample can be located without changing the orientation of the decision plane.*

Proposed Method

In this paper, we generate minority-class samples by partitioning the sequence of acoustic vectors in the enrollment utterance into a number of segments or sub-utterances, with each segment producing one GMM-supervector.

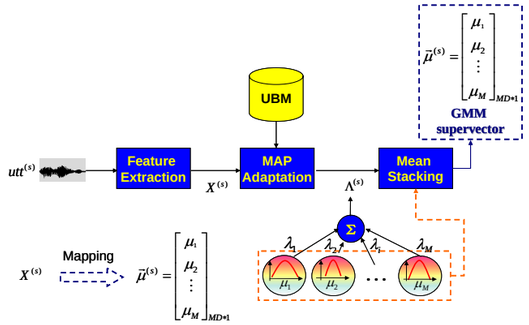

Fig.2: *Creation of GMM Supervector*

To increase the number of segments, we propose to randomize the sequence order before partitioning takes place. This randomization and partitioning process can be repeated several times to produce a desirable number of GMM-supervectors. The randomization process ensures that the GMM-supervectors are different from repetition to repetition.

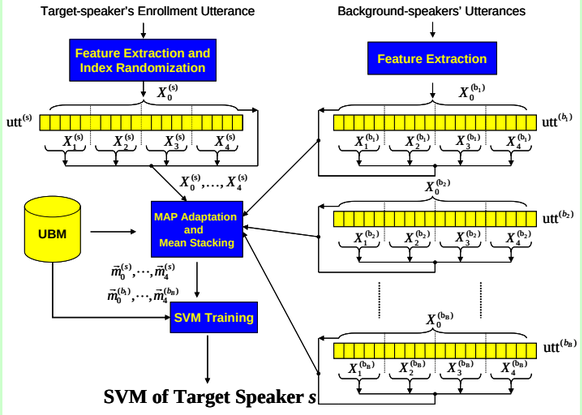

Fig.3: *Procedure of utterance partitioning with acoustic vector resampling (UP-AVR)*

Results

Fig.4 demonstrates the benefit of increasing the number of speaker-class supervectors.

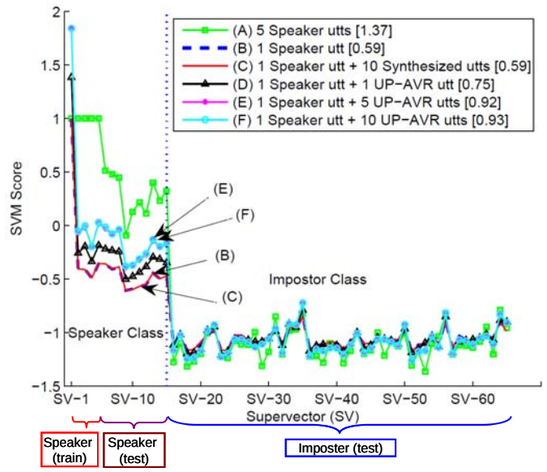

Fig.4: *Scores produced by SVMs that use one or more speaker-class supervectors (SVs) and 250 background SVs for training. The horizontal axis represents the training/testing SVs. Values inside the squared brackets are the mean difference between speaker scores and impostor scores.*

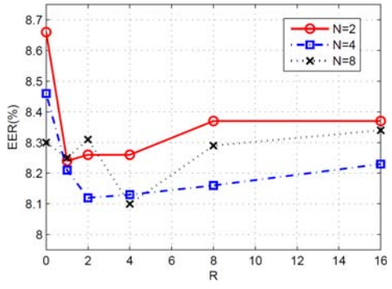

Fig.5: *Performance of UP-AVR for different numbers of partitions (N) and resampling (R) in NIST'02. When R = 0, UP-AVR is reduced to UP.*

Table1: *NIST'04*

Method	EER	MinDCF
(A) GMM-UBM+TNorm	16.05	0.0601
(B) GMM-SVM+TNorm	13.40	0.0516
(C) GMM-SVM+NAP+TNorm	10.42	0.0458
(D) GMM-SVM+NAP+TNorm+UP-AVR(5)	9.67	0.0421
(E) GMM-SVM+NAP+TNorm+UP-AVR(61)	9.63	0.0422
(F) GMM-SVM+NAP+TNorm+UP-AVR(101)	9.46	0.0419
(G) GMM-SVM+NAP+TNorm+UP-AVR(201)	9.58	0.0421

Table2: *NIST'02*

Method	EER(%)	MinDCF
(A) GMM-UBM+TNorm	10.29	0.0428
(B) GMM-SVM+NAP+TNorm	9.05	0.0362
(C) GMM-SVM+NAP+TNorm+UP(5)	8.46	0.0342
(D) GMM-SVM+NAP+TNorm+UP-AVR(33)	8.16	0.0337

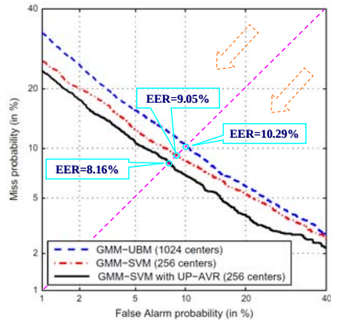

Fig.6(a) NIST'02

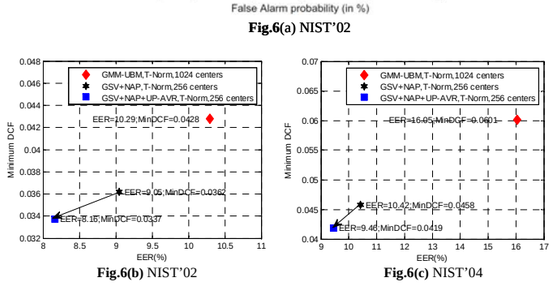

Fig.6(b) NIST'02 **Fig.6(c)** NIST'04

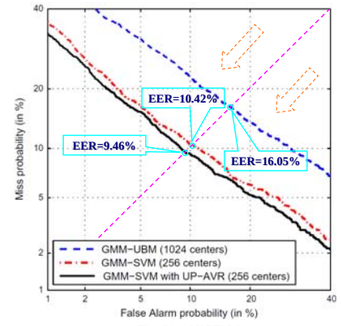

Fig.6(d) NIST'04

Fig. 6: *The DET curve and Minimum DCF versus EER*

Conclusion

A useful set of speaker-class supervectors can be generated by randomizing the sequence order of acoustic vectors in enrollment utterance for GMM-SVM speaker verification.