ValidSoft system for NIST SRE 2010

Benoit Fauve - 9 Devonshire Square, London EC2M, UK
benoit.fauve@validsoft.com

Overview

ValidSoft systems for NIST SRE 2010 are based on support vector machines (SVM) with linear kernels applied to FA modeled (*channel only*) Gaussian mixture models (GMM) supervectors.

VLD_1 : **VLD_2** on phn-phn for **VLD_3** phn-mic, mic-mic.

VLD_2 : Combination of 3 SVM-FA systems with different use of background data.

VLD_3 : SVM-FA whose channel matrix is only based on microphone recordings.

- **Frontend:**

M: MFCC **L**:LFCC

Speech activity detection (**SAD**): Tri-Gaussian on energy component. Threshold is set according to the Gaussian of highest energy, based on its weight (W) or mean (M)

- **Background data:**

Comes from NIST SRE 04, 05 and 06. To accommodate microphone data, mic recordings from SRE 05 are used.

Systems		Feature		UBM			Channel matrix	
		SAD	LFCC / MFCC	Set	# seg male	# seg female	Training Set	Rank
VLD_2	1	M	L	Phn04+Mic05	*354*	*352*	Phn04+Mic05	*80*
	2	M	M	Phn04	*219*	*196*	Phn04	*40*
	3	W	L	Phn04	*219*	*196*	Phn04	*40*
VLD_3		M	L	Phn04+Mic05	*354*	*352*	Mic05	*40*

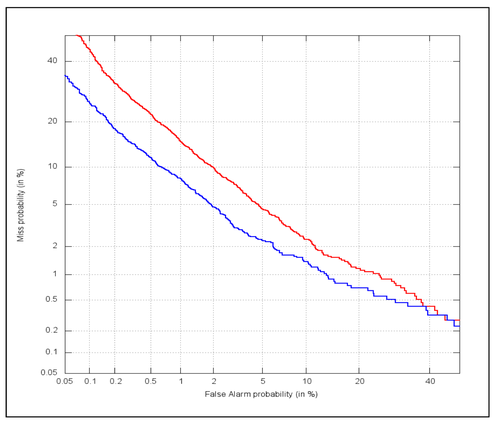
Systems		Svm cohorts			Score normalisation cohorts				
		Set	# seg male	# seg female	Set	# seg male	# seg female	T	TZ
VLD_2	1	Phn04-05-06+Mic05	*981*	*1186*	Phn04+Mic05	*250*	*275*		x
	2	Phn04-05-06+Mic05	*981*	*1186*	Phn04	*115*	*119*	x	
	3	Phn04-05-06+Mic05	*981*	*1186*	Phn04	*115*	*119*	x	
VLD_3		Phn04+Mic05-06	*822*	*883*	Phn04+Mic05	*250*	*275*		x

Results of VLD_1 primary system:

	all			*male*	*female*
	EER	*minDCF*	*minDCF old*	*EER*	*EER*
det1	4.78	0.807	0.249	4.04	5.53
det2	8.03	0.945	0.380	6.02	9.73
det3	5.79	0.667	0.238	4.42	6.29
det4	5.63	0.725	0.246	4.32	6.83
det5	5.58	0.588	0.520	5.56	5.27
det6	5.71	0.882	0.276	6.01	5.81
det7	9.02	0.745	0.349	7.57	9.91
det8	1.21	0.450	0.080	0.84	1.39
det9	3.23	0.429	0.135	2.54	3.33

More data is better data...

- Larger SVM cohorts were found to be a source of improvement.
- A lot of development time taken on usage of background data, but...



— 1 submitted (VLD_3): Channel K=40 from mic05

— 2 not run on time: same as **2**

+ concatenated 20 int08 eigenchannel

+ ~100 extra int08 samples in SVM cohort

..., the research is how to use it.

Some experiments on new threshold

- How to accommodate the low number of False Alarms? High ratio 'time spent on the question' / 'results'.
- Investigation in getting the threshold from score distribution statistics (no need to compute test on an extended data set?)

- estimation with 1 Gaussian:

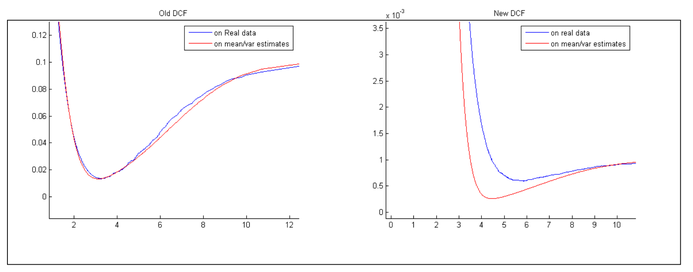
$dC(\theta)/d\theta = 0$ take a quadratic form -> 1 exact solution θ_{opt} function of C_{miss}, C_{fa}, P_{target}, m_T, s_T, m_F and s_F.

- estimation with >= 2 Gaussians (EM) -> numerical approximation of θ_{opt}

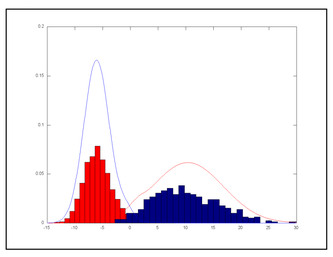
Some observations:

- Good approximation with old cost parameters, and only mean and variance estimates (mono Gaussian). This is not where the problem lies, but this shows that the idea is worth investigating.

Extended data set VS mono-Gaussian estimates



Example of tri-Gaussian approximation



- Under estimation of threshold with mono Gaussian (high score impostor are not reflected) need a finer estimate -> more Gaussians.
- Modelling with more Gaussians: better estimate of threshold (as compared with estimation on extended data set).

More analysis is needed.

Conclusions

Introduction of new DCF parameters was an interesting problem but a distraction for other development.

What has worked:

- More background data in the system (in particular in channel matrices and SVM cohort)