alpineon

Naložba v vašo prihodnost
OPERACIJO DELNO FINANCIRA EVROPSKA UNIJA
Evropski socialni sklad

ALPineon System for NIST SRE 2010

Boštjan Vesnicer

bostjan@alpineon.si

Brno, June 2010

Overview

- Core-core condition
- Focus on telephone speech
- Joint Factor Analysis
- Score fusion (Likelihood ratio, SVM)

Configuration

- Silence removal based on NIST's transcriptions
- Unnormalized features: MFCCs (60)
- Gender-independent UBM (2048 comp.)
- Gender-independent JFA (300 speaker factors, 100 channel factors)
- No score normalization

Data

	UBM	JFA	SVM background	DEV
2004	●	●	●	
2005	●	●	●	
2006	●	●	●	
2008				●

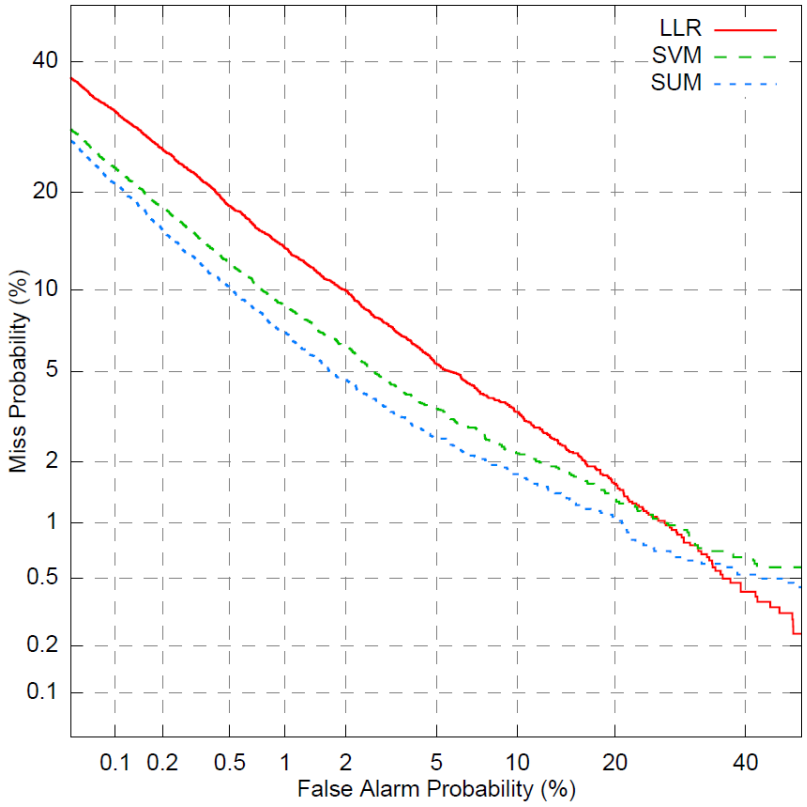
- JFA hyperparameters were estimated in a decoupled fashion (Kenny, 2008)

Scoring

- Likelihood $P(x \mid s) = \int P(x \mid s, c) P(c) dc$
  - s: point estimate of the target speaker supervector
  - P(c): channel prior
- SVM
  - JFA speaker supervectors
  - Rank normalization
  - Linear kernel
- No score normalization was performed

Development results

- NIST SRE 2008 1conv4w-1conv4w (tel. speech)



	EER (%)	DCF (x10^4)
LLR	5.2	6.5
SVM	4.0	5.3
SUM (LLR + SVM)	3.4	5.1

alpineon

Conclusion

- Promising results on the telephone data despite
  - Gender-independent modeling (UBM, JFA, ...)
  - Unnormalized features
  - Unnormalized scores: score normalization is not always effective, e.g. (Kenny, 2010)
- The same is not true for the microphone data
  - It seems that feature normalization is crucial for the microphone data (at least if UBM was trained on the tel. speech)

References

- Kenny, P., Ouellet, P., Dehak, N., Gupta, V., and Dumouchel, P., A Study of Inter-Speaker Variability in Speaker Verification, IEEE Transactions on Audio, Speech and Language Processing, July 2008.
- Kenny, P., Bayesian Speaker Verification with Heavy-Tailed Priors, Odyssey Speaker and Language Recognition Workshop, Brno, Czech Republic, June 2010.