Compensation of speech codec in speaker recognition

April 14, 2022

Speaker recognition

- Speaker recognition tasks: speaker verification, speaker identification.
- Usage: electronic commerce, electronic banking transactions, forensic investigations.
- Problem: session variability. Codecs contribute to session variability.

Feature generation

MFCC got from signal: $x_1, \ldots, x_n$. Overlapping fragments (e.g. of length 20 ms, shift 10 ms). Each fragment maps to feature vector.

Feature generation

GMM supervector *M* [4].

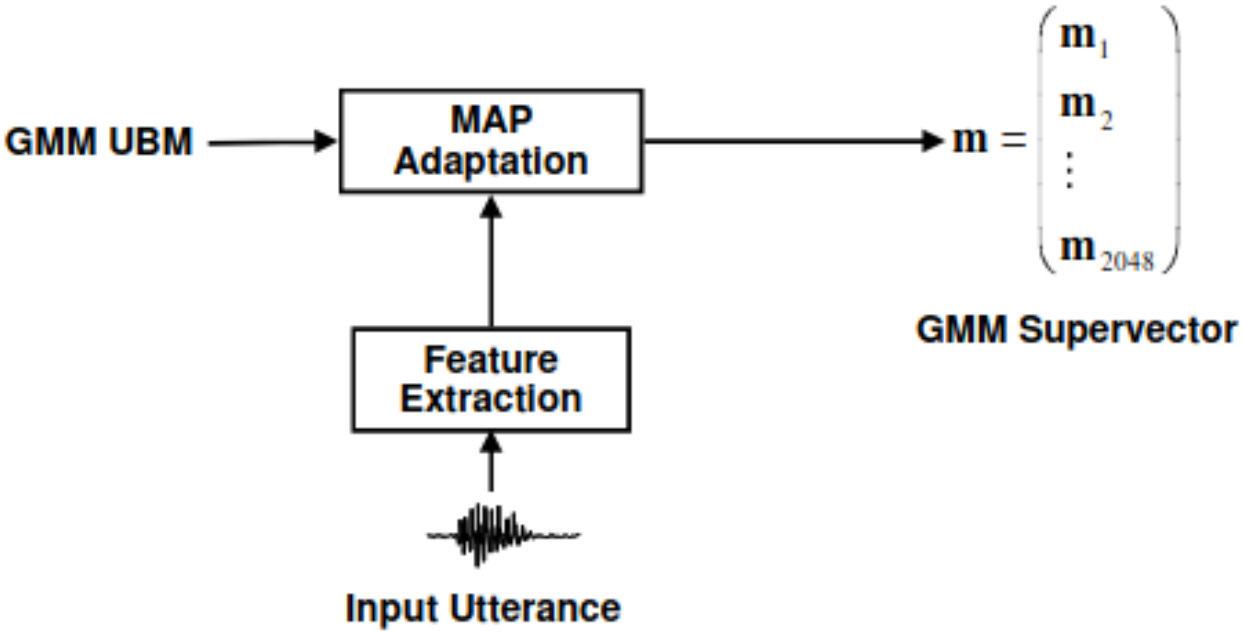


Fig. 1. GMM supervector concept

Joint factor analysis

- Classical approach. Speaker utterance is represented by a supervector

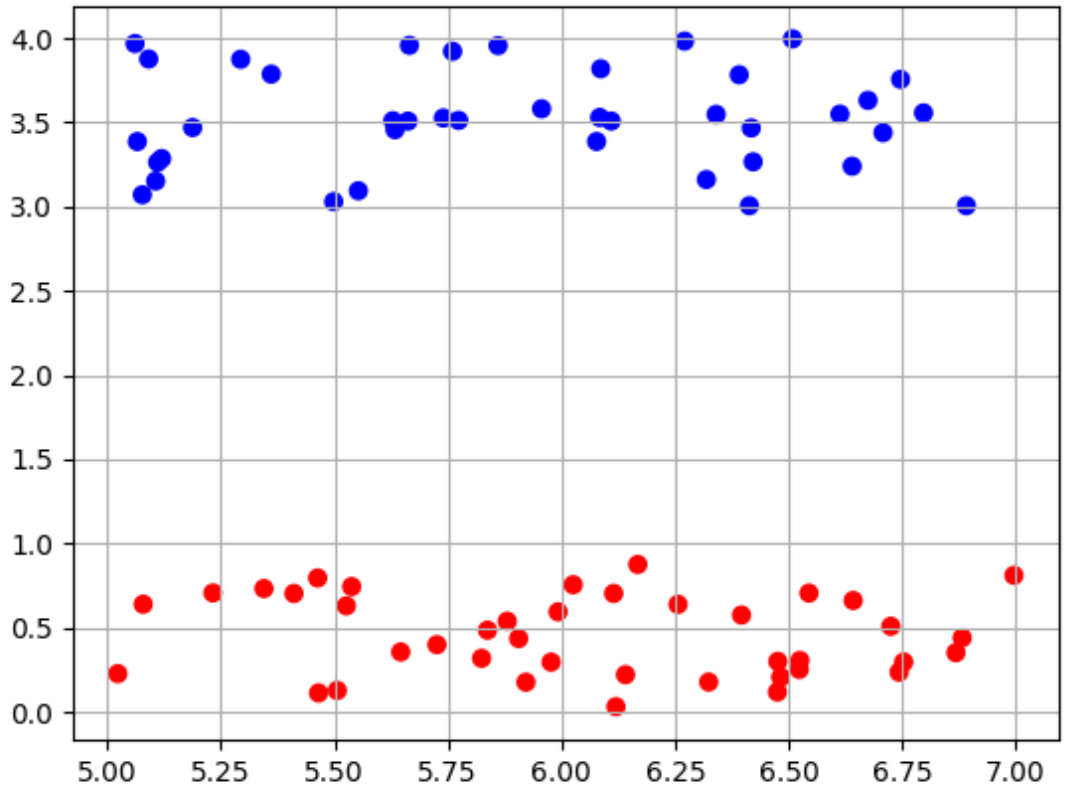
$$M = m + Ux + Vy.$$

  V defines speaker subspace, U defines session subspace.
- Drawback: channel factors contain information about the speaker.
- Total variability space:

$$M = m + Tw.$$

Session variability compensation

- NAP — Nuissance Attribute Projection.
- WCCN — Within Class Covariance Normalization.
- LDA — Linear Discriminant Analysis: find features that discriminate speakers best.



Codecs

- Codecs by sampling frequency: NB — narrow bandwidth (sampling frequency 8kHz), WB — broad bandwidth (sampling frequency 16 kHz).
- Codecs by bit rate: low or high bit rate.
- Codecs with higher bit rate affect speaker recognition less.
- NB codecs — worse recognition effectiveness.
- Some frequency bands more discriminative than others [3].

Experiments

- Data transmitted through a digital telephone network.
- First experiment: MOS quality measurement for different codecs. t
- Second experiment: influence of speech codec on speaker identification. 4 speakers, 7 codecs. High bit-rate codecs provide better speech quality and preserve better speaker discriminative capacity.
- Third experiment: NIST dataset, use of S-LDA variability compensation. Codecs G.722, GSM 6.20. S-LDA method improves speaker recognition.

Literature I


José R. Calvo de Lara, Flavio J. Reyes Diaz, Gabriel Hernández Sierra, Orlando Jimenez Alcazar.
A method to compensate the influence of speech codec in speaker recognition.
International Journal of Speech Technology, 2018.


Najim Dehak et.al.
Front-End Factor Analysis for Speaker Verification.
IEEE Transactions on Audio Speech and Language Processing, 2011.


Fernández L., Wagner M., Möller S.
Advantages of wideband over narrowband channels for speaker verification employing MFCCs and LFCCs.
ISCA Interspeech Conference Proceedings, 2014.

Literature II

W. M. Campbell, D. E. Sturim, D. A. Reynolds
Support Vector Machines using GMM Supervectors for Speaker Verification
IEEE Signal Processing Letters, 2006.