FULL-COVARIANCE UBM AND HEAVY-TAILED PLDA IN I-VECTOR SPEAKER VERIFICATION

Pavel Matejka¹, Ondrej Glembek¹, Fabio Castaldo², M.J. Alam³, ⁴, Oldrich Plchot¹, Patrick Kenny³, Lukas Burget¹, Jan Cernocky¹
¹ Brno University of Technology, Czech Republic, ² Centre de Recherche Informatique de Montreal (CRIM), Montreal, Canada, ³ Loquendo, Italy, ⁴ INRS-EMT, Montreal, Canada

• Single best system in post-analysis of ABC (Agnito+But+CRIM) NIST SRE 2010 submission was Full covariance UBM with the state-of-the-art scheme - iVector + PLDA
• Do we really need full-covariance matrices?
• Let us take a look at some analysis.

iVector + PLDA
• iVector extractor – model similar to JFA, where GMM mean supervector
  \[ \mu = m + T i \]
is constrained to leave in single subspace \( T \) spanning both speaker and channel variability \( m \) no need for speaker labels to train \( T \)
• iVector – point estimate of \( i \) can now be extracted for every recording as its low-dimensional, fixed-length representation (typically 400 dimensions)
• contains information about both speaker and channel
• are assumed to be normal distributed
• Natural choice is simplified JFA model with only single Gaussian. Such model is known as PLDA and is described by familiar equation:
  \[ i = m + V y + U x + \epsilon \]
• PLDA has nice interpretation in face verification where it was introduced by Simon J.D. Prince
• Each face image \( i \) can be constructed by adding
  • mean face \( m \)
  • linear combination of basis \( V \) corresponding to between-individual variability (moving from \( m \) in these directions gives us images that look like different people)
  • linear combination of basis \( U \) corresponding to within-individual variability (moving from \( m \) in these directions gives us images that looks like different pictures of the same person)
  • residual noise vector \( \epsilon \)
• Gaussian PLDA – assume standard normal prior for iVectors
• Heavy tailed PLDA – assume Student’s t distribution prior for iVectors

Motivations for full covariance GMM:
• Better description of feature space while preserving reasonable size of GMM mean supervector
• Higher computational complexity \( \bar{\text{bar}} \) investigation into possible simplifications
• Full covariance Gaussians are more sensitive to very low values of off-diagonal elements \( \rightarrow \) variance flooring:

\[
\begin{align*}
N_i & = \sum_i \gamma_i \xi_i (\mu - m)^T \Sigma (\mu - m) \\
T_i & = \sum_i \gamma_i \xi_i (\mu - m)
\end{align*}
\]

Different statistic normalization

Zero order statistics:
\[ f^{(0)} = \sum \gamma_i \xi_i 0 \]
First order statistics:
\[ f^{(1)} = \sum \gamma_i \xi_i \mu \]
Normalization:
\[ f^{(c)} + \sum \gamma_i \xi_i (\mu - m)^T \Sigma (\mu - m) \]

Experimental Setup
Features: MFCC 19-E, Delta + double delta
Short time cepstral mean and variance normalization over 300 frames,
Dataset: NIST SRE 2010, Extended core condition 5 – tel-tel, Female only

Historical way to iVector + PLDA

Different scoring

Amount of training data
• Full covariance | Diagonal cov. | Diagonal cov + HLDA
• iVector 400, LDA 150, Norm2, Gaussian PLDA
• big = NIST SRE 2004 + 2005 = 310 hours
• small = 3 hours subset of big set

CONCLUSION
• Full covariance UBM gives the best results
• With unity length normalization of iVector you can use Gauss PLDA
• Diagonal covariance UBM with MLLT/HLDA goes very close and have benefit of fast evaluation of Gaussians