I-vector representation based on GMM and DNN

2019.04.17

齐诏娣

Outline

- Introduction
- I-vector based GMM
- I-vector based DNN
- Conclusion

Introduction

- JFA intuition
  - A supervector for a speaker should be decomposable into speaker independent, speaker dependent, channel dependent, and residual components
  - A given speaker GMM supervector s can be decomposed as follows:

$$s = m + Vy + Ux + Dz$$

"Ideal" speaker supervector

Speaker-independent component

Speaker-dependent component

Channel-dependent component

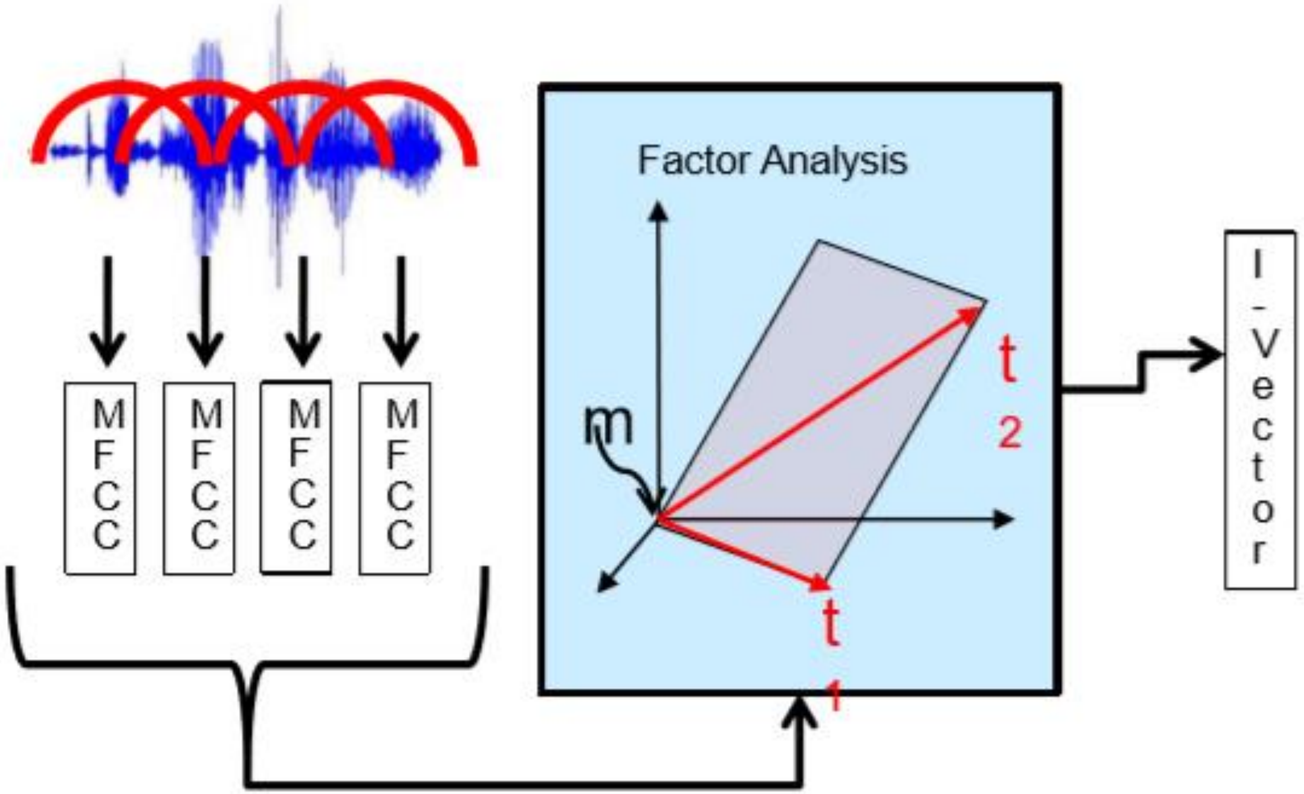
Speaker-dependent residual component

- JFA的初衷是移除说话人均值超矢量在本征信道空间的影响，让y（说话人空间）具有很好的抗信道失配能力
- 但，信道条件本身很难用数学公式来明确界定它，导致可能X（信道空间）包含说话人信息
- 论文[2]中，也证明了信道因子包含了说话人信息。
- 且，JFA因子分析训练语料要求高，计算复杂

I-vector based GMM

- An i-vector system uses a set of low-dimensional total variability factors (w) to represent each conversation side. Each factor controls an eigen-dimension of the total variability matrix (T), and are known as the i-vectors.

$$\underbrace{s}_{\text{Conversation side supervector}} = m + \underbrace{T}_{\text{Total-variability matrix}}\underbrace{w}_{\text{i-vector}}$$



MFCC
MFCC
MFCC
MFCC
Factor Analysis
m
t2
t1
I-Vector

- 计算流程
  - 1. 计算Baum-Welch统计量：

$$N_c = \sum_{t=1}^{L} \mathrm{P}\left(c|y_t, \Omega\right)$$

$$F_c = \sum_{t=1}^{L} \mathrm{P}\left(c|y_t, \Omega\right) y_t,$$

    - 在Baum-Welch统计量中，定义一阶统计量

$$\tilde{F}_c = \sum_{t=1}^{L} \mathrm{P}\left(c|y_t, \Omega\right)(y_t - m_c),$$

- 2. EM算法：T
  - E步：
    - 计算隐变量w的后验分布
    - $p(\mathbf{w}|\mathcal{X}) = \mathcal{N}(\mathbf{w}|\mathbf{L}^{-1}\mathbf{T}^T\mathbf{\Sigma}^{-1}\mathbf{F}, \mathbf{L}^{-1})$ 其中，$\mathbf{L} = \mathbf{I} + \mathbf{T}^T\mathbf{\Sigma}^{-1}\mathbf{N}\mathbf{T}$
  - M步：
    - 更新全局差异空间矩阵T
- 3. i-vector 提取：w

$$w = (I + T^t\Sigma^{-1}N(u)T)^{-1}.T^t\Sigma^{-1}\tilde{F}(u).$$

I-vector based DNN

- 基本思想
  - 利用基于深度神经网络的语音识别模型DNN-ASR替换通过非监督聚类的高斯混合模型UBM。
- 特点
  - DNN能从大量样本中学习到高度抽象的音素特征，并对噪声有很强的免疫力。

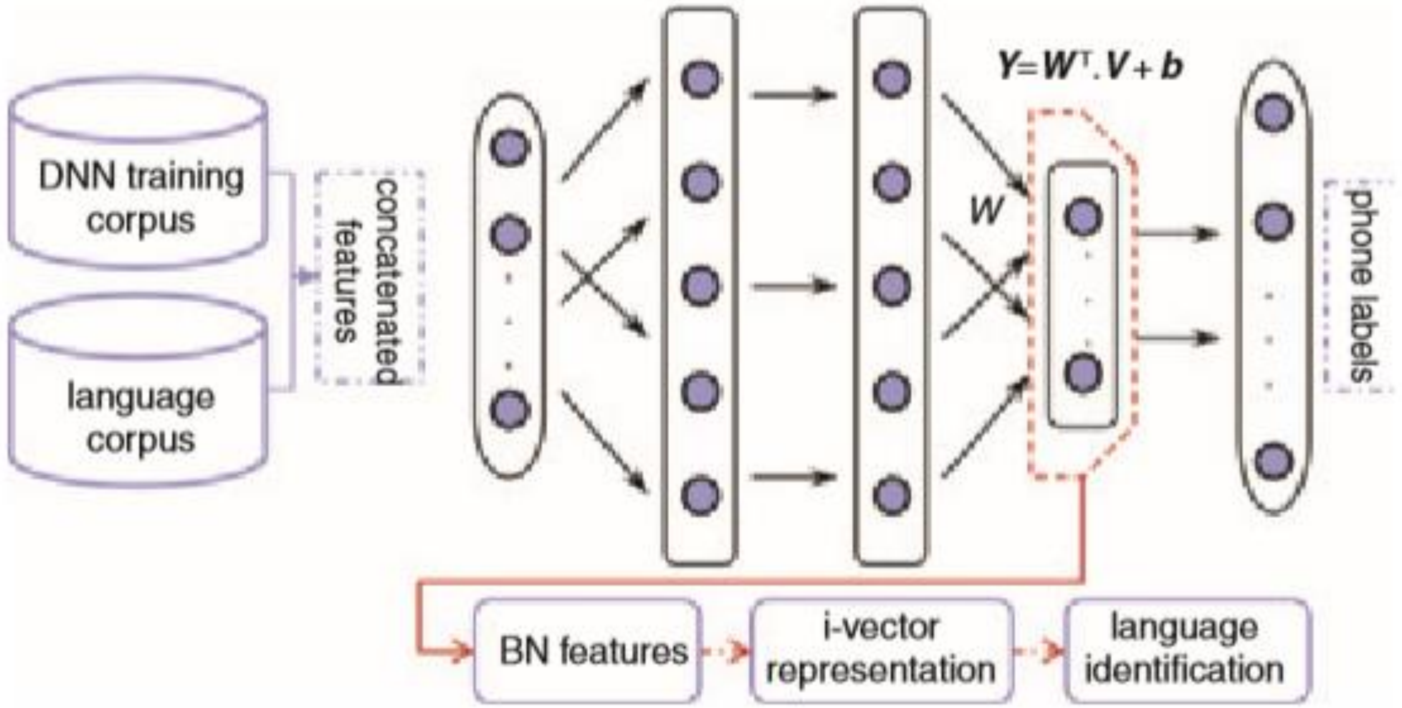


Fig 2 Proposed i-vector representation based on BN features for automatic LID [3]

Conclusion

- GMM i-vector & DNN i-vector不同
  - 用于计算充分统计量的后验概率不同
  - 训练方式不同
- I-vector 中T矩阵同时对说话人和信道两个空间建模，w也带有信道信息，所以后续应做i-vector的信道补偿工作
  - 在[1]中提到的有3种（1）类内协方差归一化WCCN（2）线性判别分析LDA（3）扰动属性投影 NAP
  - 之后，PLDA

Reference

- [1] Front-End Factor Analysis for Speaker Verification
- [2] Discriminative and generative approaches for long and shortterm speaker characteristic modeling application to speaker verification
- [3] i-vector representation based on bottleneck features for language identification