Non-Iterative Parameter Estimation for Total Variability Model Using Randomized SVD

Ruchir Travadi, Shrikanth Narayanan
Signal Analysis and Interpretation Laboratory (SAIL), Viterbi School of Engineering, University of Southern California, Los Angeles, CA 90089
travadi@usc.edu

Total Variability Model

Parametric model for explaining changes in feature space distribution across utterances
Used in tasks like speaker and language identification

Model Formulation

Feature vectors : $\mathbf{X}_u = \{\boldsymbol{x}_{ut}\}_{t=1}^{T_u}$, **ivector** : $\boldsymbol{w}_u$

$$\boldsymbol{W}_u \sim \mathcal{N}(\mathbf{0}, \mathbf{I}), \quad \boldsymbol{x}_{ut} | \boldsymbol{w}_u \sim \text{GMM} \left\{ p_c,\ \boldsymbol{\mu}_{uc} = \boldsymbol{\mu}_c + \mathbf{T}_c \boldsymbol{w}_u,\ \Sigma_c \right\}_{c=1}^{C}$$

$$M_0 = \begin{bmatrix} \boldsymbol{\mu}_1 \\ \vdots \\ \boldsymbol{\mu}_C \end{bmatrix}, \quad M_u = \begin{bmatrix} \boldsymbol{\mu}_{u1} \\ \vdots \\ \boldsymbol{\mu}_{uC} \end{bmatrix}, \quad \mathbf{T} = \begin{bmatrix} \mathbf{T}_1 \\ \vdots \\ \mathbf{T}_C \end{bmatrix} \qquad M_u = M_0 + \mathbf{T}\boldsymbol{w}_u$$

Conventional Training Method

Expectation Maximization

Expectation : Evaluate the posterior distribution for ivectors

$$\boldsymbol{w}_u | \mathbf{X}_u, \Theta^{(n-1)} \sim \mathcal{N}\left(\left(\mathbf{I} + \mathbf{T}^\mathsf{T}\Sigma^{-1}\mathbf{N}_u\mathbf{T}\right)^{-1} \mathbf{T}^\mathsf{T}\Sigma^{-1}\boldsymbol{F}_u, \left(\mathbf{I} + \mathbf{T}^\mathsf{T}\Sigma^{-1}\mathbf{N}_u\mathbf{T}\right)^{-1} \right)$$

Maximization : Maximize the following objective function

$$\mathcal{L}(\Theta) = \sum_{u=1}^{U} \mathbb{E}_{\boldsymbol{w}_u | \mathbf{X}_u, \Theta^{(n-1)}} \left[\log P(\mathbf{X}_u, \boldsymbol{w}_u | \mathbf{C}_u, \Theta^{(n)}) \right]$$

Expensive to evaluate all ivectors every iteration

Key Result

Assuming

$$\boldsymbol{N_{uc}} \approx \boldsymbol{T_u p_c}, \quad \mathbf{T}^\mathsf{T}\Sigma^{-1}\mathbf{P}\mathbf{T} + \tfrac{1}{T_u}\mathbf{I} \approx \mathbf{T}^\mathsf{T}\Sigma^{-1}\mathbf{P}\mathbf{T} + \tfrac{1}{T}\mathbf{I}$$

Given

Covariance matrices Σ_c, UBM posteriors $\mathbf{C}_u$

Objective Function : Marginalized Likelihood

$$\sum_{u=1}^{U} \log P(\mathbf{X}_u | \mathbf{C}_u, \Theta) = \sum_{u=1}^{U} \log \left(\int_{\boldsymbol{w}_u} P(\mathbf{X}_u, \boldsymbol{w}_u | \mathbf{C}_u, \Theta) \right)$$

Is maximized for $\mathbf{T}$ = Partial SVD of normalized statistics

Proposed Training Method

Randomized SVD of Normalized Statistics

$$\Sigma_c^{-1} = \mathbf{L}_c \mathbf{L}_c^\mathsf{T}, \quad \widetilde{F}_u = \begin{bmatrix} \frac{1}{\sqrt{N_{u1}}} \mathbf{L}_1{}^\mathsf{T} \boldsymbol{F}_{u1} \\ \vdots \\ \frac{1}{\sqrt{N_{uC}}} \mathbf{L}_\mathbf{C}{}^\mathsf{T} \boldsymbol{F}_{uC} \end{bmatrix}, \quad \widetilde{\mathbf{F}} = \begin{bmatrix} \widetilde{F}_1 & \dots & \widetilde{F}_U \end{bmatrix}$$

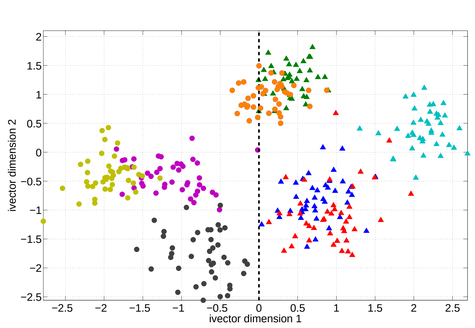
$$\widetilde{\mathbf{F}} = \mathbf{U}_F \mathbf{D}_F \mathbf{V}_F^\mathsf{T}, \quad \widetilde{\mathbf{T}} = \mathbf{U}_F \mathbf{D}_F, \quad \mathbf{T} = \Sigma \mathbf{L} \mathbf{P}^{-\frac{1}{2}} \widetilde{\mathbf{T}}$$

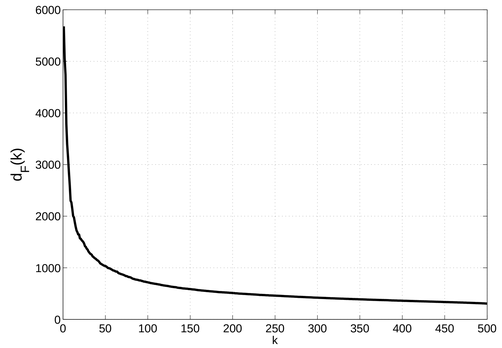
Key Advantages

Non-iterative, Efficient, Parallelizable, Interpretable

No need for iterations like in EM

i-vector dimensions ranked by singular values, no re-training for different i-vector dimensionality





Experimental Results

RATS language identification data : 6 languages, 16000 utterances per language, 30s each

	EM		**SVD-1**		**SVD-2**	
T_θ	33.5 hrs		**30 min**		**30 min**	
T_w	1.13 sec		**0.12 sec**		1.13 sec	
Dur	*EER*	*ACC*	*EER*	*ACC*	*EER*	*ACC*
10	**8.30**	83.50	8.78	**83.60**	9.26	81.15
5	**10.04**	**80.80**	10.40	80.55	10.94	78.05
3	**13.41**	72.75	13.59	**74.25**	14.97	70.55
1	**21.59**	56.45	22.13	**58.35**	22.61	54.65