Joint Bayesian Gaussian Discriminant Analysis For Speaker Verification

Yiyan Wang wangyiya14@mails.tsinghua.edu.cn

Haotian Xu xht13@mails.Tsinghua.edu.cn

Zhijian Ou ozj@tsinghua.edu.cn

Speech Processing and Machine Intelligence (SPMI) Lab, Tsinghua University, Beijing, China

Joint Bayesian (JB) Model for Speaker Verification

The j-th i-vector of speaker i, denoted by $x_{ij} \in R^d$, is decomposed as:

$$\underbrace{\boldsymbol{\mu_i}}_{\text{Speaker identity variable}} + \boldsymbol{\varepsilon_{ij}} \rightarrow \text{Within-speaker variability}$$

$$\boldsymbol{x_{ij}} = \boldsymbol{\mu_i} + \boldsymbol{\varepsilon_{ij}}$$

- Two independent Gaussians: $\mu_i \sim N(0, S_\mu)$ $\varepsilon_{ij} \sim N(0, S_\varepsilon)$
- Training (**EM algorithm**)

$$max_\Theta \sum_i E_{p(h_i|x_i;\Theta^t)}[\log p(h_i; \Theta^{t+1})]$$

- Testing

$$r(x_1, x_2) = log \frac{p(x_1, x_2|H_I)}{p(x_1, x_2|H_E)}$$
$$= logp(x_1, x_2) - logp(x_1) - logp(x_2)$$

Efficient testing: Simultaneous Diagonalization (SD)

Testing: do simultaneous diagonalization of S_μ and S_ε

$$\Phi^T S_\mu \Phi = K$$
$$\Phi^T S_\varepsilon \Phi = I$$

Diagonal matrix

- Define $\Psi = \Phi^{-T} \longrightarrow S_\mu = \Psi K \Psi^T \quad S_\varepsilon = \Psi I \Psi^T$

$$\Sigma_{x_i} = \begin{bmatrix} S_\mu + S_\varepsilon & S_\mu & \cdots & S_\mu \\ S_\mu & S_\mu + S_\varepsilon & \cdots & S_\mu \\ \vdots & \vdots & \ddots & \vdots \\ S_\mu & S_\mu & S_\mu & S_\mu + S_\varepsilon \end{bmatrix} = \Omega \begin{bmatrix} K+I & K & \cdots & K \\ K & K+I & \cdots & K \\ \vdots & \vdots & \ddots & \vdots \\ K & K & K & K+I \end{bmatrix} \Omega^T$$

where $\Omega = diag(\Psi; \cdots; \Psi)$

- The calculation of $p(x_i)$ could be accelerated, which only involves inversion of diagonal matrices.

Complexity: $O(d^3) \rightarrow O(d)$

Connection with PLDAs

Method	JB	two-covariance	SPLDA	Kaldi PLDA
Observation	$x_i = \{x_{ij}, j = 1, \ldots, m_i\}$			$\bar{x}_i = \frac{1}{m_i}\sum_{j=1}^{m_i} x_{ij}$
Model	$x_{ij} = \mu_i + \varepsilon_{ij}$		$x_{ij} = Fz_i + \varepsilon_{ij}$	$\bar{x}_i = \mu_i + \varepsilon_{i1}$
h_i	$\{\mu_i, \{\varepsilon_{ij}\}\}$	$\{\mu_i\}$	$\{z_i\}$	$\{\mu_i, \varepsilon_{i1}\}$
EM objective function $Q(\Theta_t, \Theta_{t+1})$	$E_{p(h_i\|x_i)}[logp(h_i)]$	$E_{p(h_i\|x_i)}[logp(x_i, h_i)]$		$E_{p(h_i\|\bar{x}_i)}[logp(h_i)]$
Subspace dimensionality setting	loose		strict	loose
EM convergence	fast	slow		fast

Table 1. The summary of the similarities and difference between JB, SPLDA, Kaldi PLDA and the two-covariance model, x_{ij} denotes the j-th i-vector of speaker i. $\mu_i \sim N(0, S_\mu)$ is the identity variable for speaker I, modeled by the between-class covariance S_μ, $\varepsilon_{ij} \sim N(0, S_\varepsilon)$ is the intersession residual, modeled by the within-class covariance S_ε. For SPLDA, $z_i \sim N(0, I)$ stands for the identity variable.

- **EM algorithm for SPLDA:**

$$max_\Theta \Sigma_i E_{p(z_i|x_i;\,\Theta_t)} logp(x_i, z_i; \Theta_{t+1}) \longleftrightarrow min_\Theta \Sigma_i \Sigma_j trace(\Lambda_{t+1}^{-1} E[(x_{ij} - F_{t+1}z_i)(x_{ij} - F_{t+1}z_i)^T])$$

$$E[z_i] = F_t^T (F_t F_t^T + \Lambda_t)^{-1} x_{ij} \downarrow$$

When Λ_t is small and $F_{t+1} \approx F_t$, $\quad x_{ij} - F_{t+1} \cdot E[z_i] \approx x_{ij} - F_{t+1} \cdot F_t^T (F_t F_t^T)^{-1} x_{ij} = 0$

The EM update for SPLDA could easily be **stuck into non-local minima** with small Λ_t.

The EM update for JB does not have such problem.

JB calculates the joint likelihood

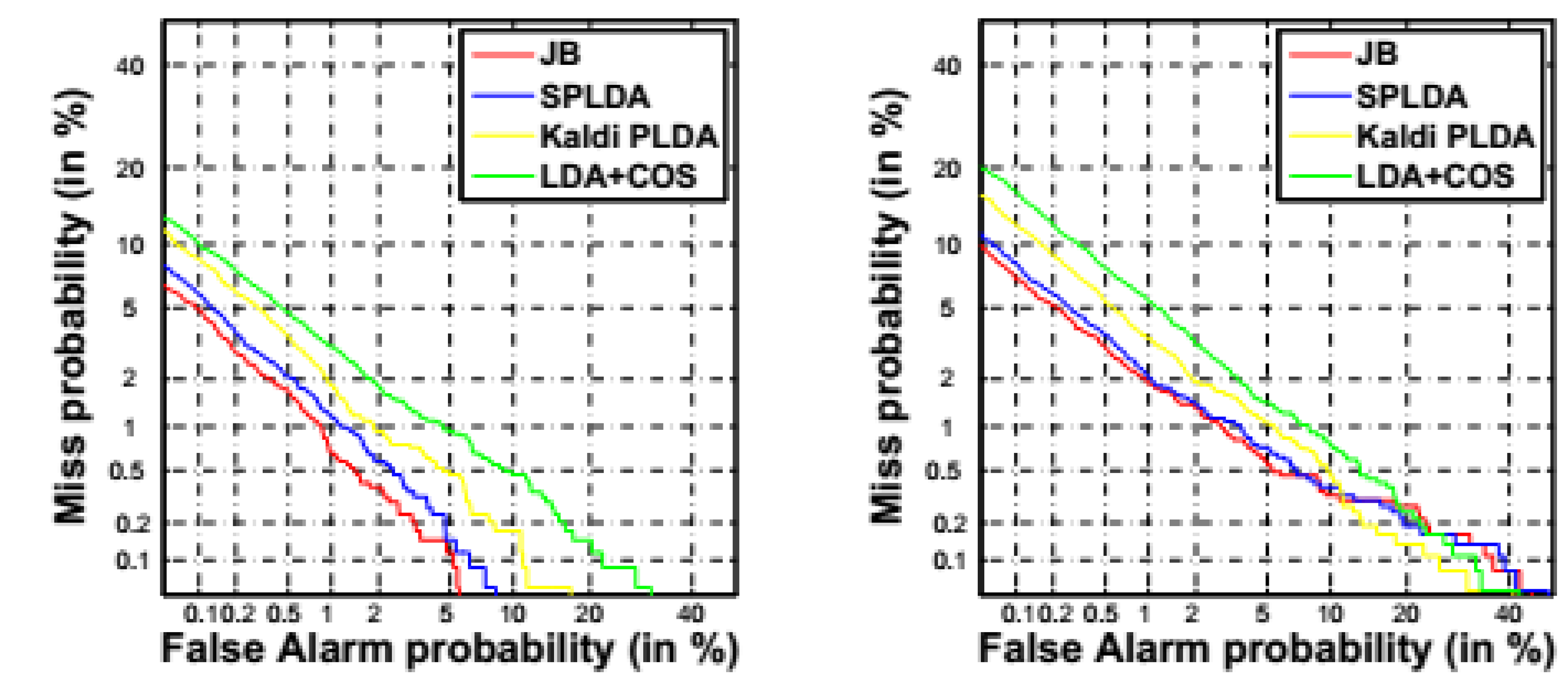
$$p(x_i) = N(0, \Sigma_{x_i})$$

Kaldi calculates the likelihood of the single average i-vector $\bar{x_i}$.

$$p(\bar{x_i}) = N\left(0, FF^T + \frac{1}{m_i}\Lambda\right)$$

Experiments

- **Speaker Verification Performance**



(a) SRE10 MALE (b) SRE10 FEMALE

Fig. 1. DET curves in SRE10 core condition 5 evaluation.

System	SRE10 MALE			SRE10 FEMALE		
	EER	DCF10	DCF08	EER	DCF10	DCF08
LDA+COS	1.905	0.292	0.091	2.619	0.399	0.126
Kaldi PLDA	1.299	0.284	0.079	1.944	0.345	0.102
SPLDA	1.010	0.217	0.055	1.621	0.287	0.079
JB	**0.894**	**0.188**	**0.048**	**1.485**	**0.245**	**0.069**

- **Subspace Dimensionality**

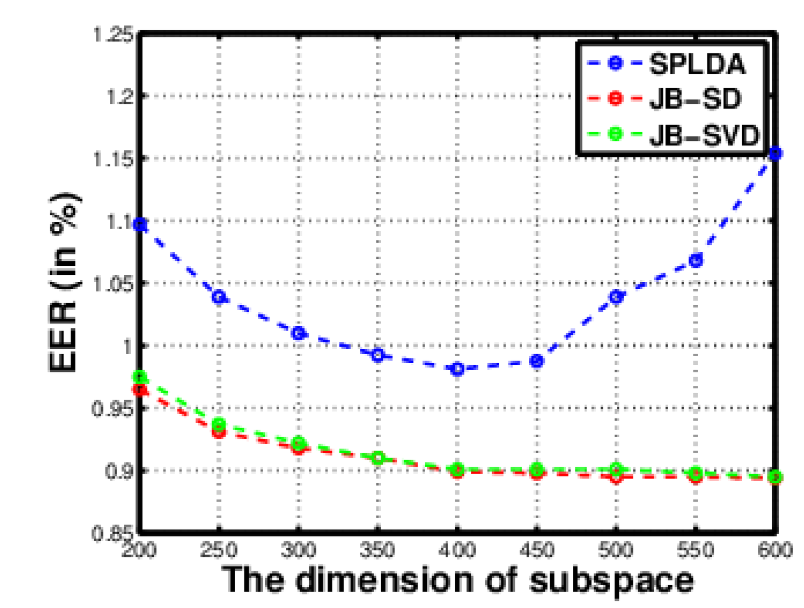
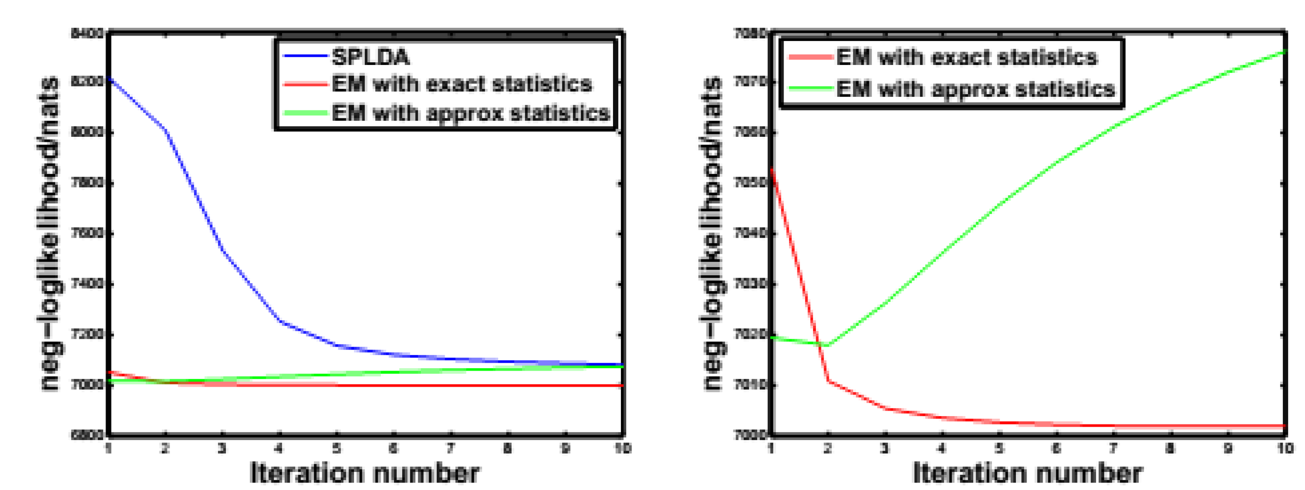


Fig. 2. The influence of subspace dimensionality on JB and SPLDA using NIST SRE10 core condition male test data.

- **Convergence Rate**



(a) (b)

Fig. 3. (a) The negative log-likelihood of JB (EM with exact or approximated statistics) and SPLDA during training. (b) The zoom-in of negative log-likelihood convergence curves for JB with exact and approximated EM statistics.