Deep Neural Network Driven Mixture of PLDA for Robust I-vector Speaker Verification

Na LI[1], Man-Wai MAK[1], and Jen-Tzung Chien[2]

[1]Dept. of Electronic and Information Engineering, The Hong Kong Polytechnic University

[2]Dept. of Electrical and Computer Engineering, National Chiao Tung University

Introduction

- Noise-level variability causes the i-vectors to form heterogeneous clusters, and i-vectors with similar SNRs tend to cluster together.
- This phenomenon indicates that SNR information can be used to guide the training of PLDA mixture models.
- This paper proposes using an SNR-aware DNN to guide the training of PLDA mixture models, resulting in a more reasonable soft division of the i-vector space.
- Experiments on NIST 2012 SRE demonstrate the effectiveness of the proposed framework compared with PLDA and conventional mixture of PLDA.

Background

SNR-independent mixture of PLDA (SI-mPLDA):

$$p(\mathbf{x}_{ij}) = \sum_{k=1}^{K} \varphi_k N(\mathbf{x}_{ij} \mid \mathbf{m}_k, \mathbf{V}_k\mathbf{V}_k^{\mathsf{T}} + \mathbf{\Sigma}_k)$$

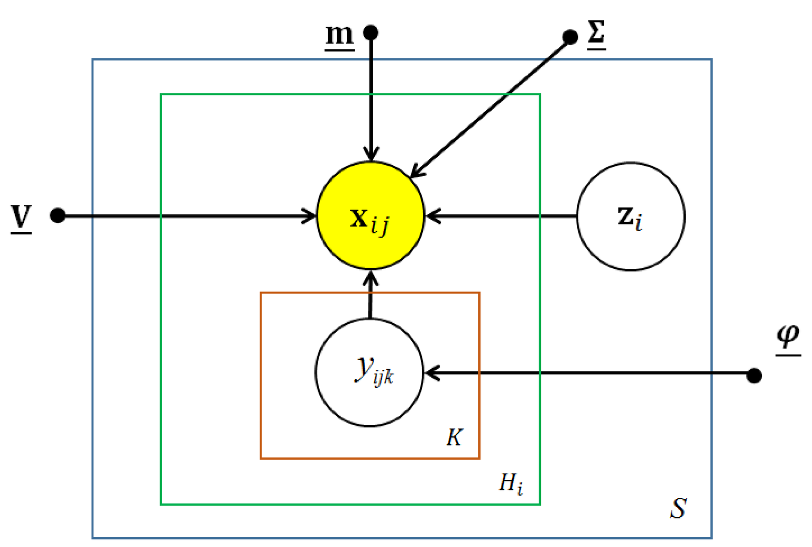

Fig.1: *Graphical model of SI-mPLDA.*

SNR-dependent mixture of PLDA (SD-mPLDA):

$$p(\mathbf{x}_{ij}, \ell_{ij}) = p(\ell_{ij}) \sum_{k=1}^{K} \gamma_\ell(y_{ijk}) N(\mathbf{x}_{ij} \mid \mathbf{m}_k, \mathbf{V}_k\mathbf{V}_k^{\mathsf{T}} + \mathbf{\Sigma}_k)$$

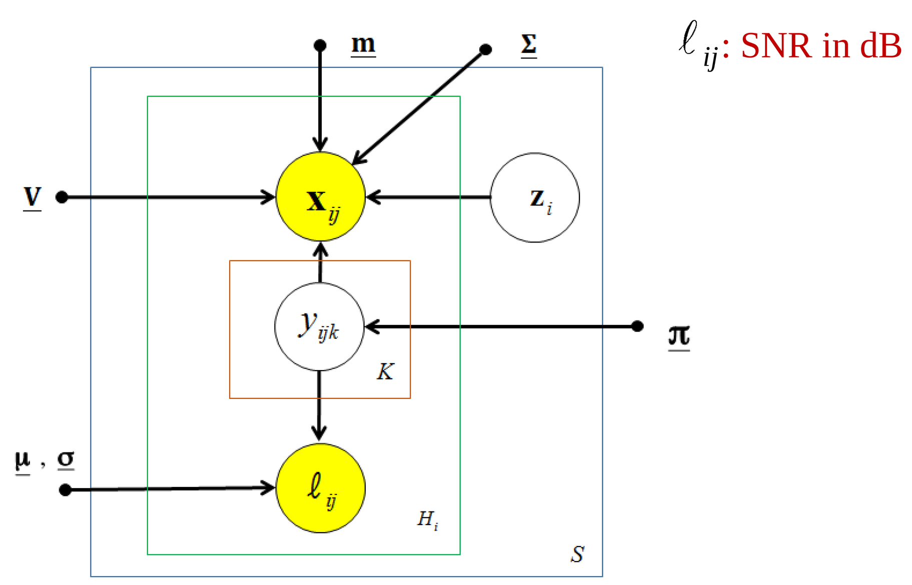

Fig.2: *Graphical model of SD-mPLDA.*

Proposed Method

The SNR subgroup posteriors generated from an SNR-aware DNN are used as the posteriors of the indicator variables in the mixture of PLDA to guide the training of the mixture model.

SNR Subgroups:

The training set is divided into different SNR groups according to the SNRs of the training utterances.

- **Table1:** *SNR ranges in dB for different numbers of SNR groups (K)*

K	Group 1	Group 2	Group 3	Group 4	Group 5
2	(-∞,20]	(20,∞)	-	-	-
3	(-∞,8]	(8, 20]	(20, ∞)	-	-
4	(-∞,8]	(8, 14]	(14,20]	(20, ∞)	-
5	(-∞,4]	(4,8]	(8,14]	(14,20]	(20, ∞)

SNR-aware DNN:

Posterior probability of the *k*-th SNR group is:

$$\gamma_{\mathbf{x}_{ij}}(y_{ijk}) \equiv P(y_{ijk} = 1 \mid \mathbf{x}_{ij}, \mathbf{w})$$

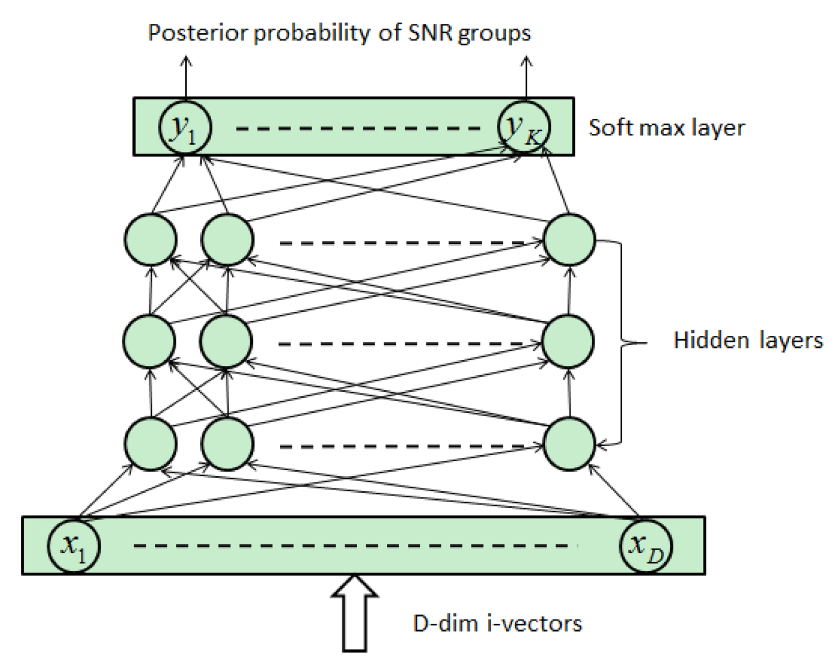

Fig.3: *Schematic diagram of the SNR-aware DNN.*

DNN-driven mixture of PLDA (DNN-mPLDA):

$$p(\mathbf{x}_{ij}) = \sum_{k=1}^{K} \gamma_{\mathbf{x}_{ij}}(y_{ijk}) N(\mathbf{x}_{ij} \mid \mathbf{m}_k, \mathbf{V}_k\mathbf{V}_k^{\mathsf{T}} + \mathbf{\Sigma}_k)$$

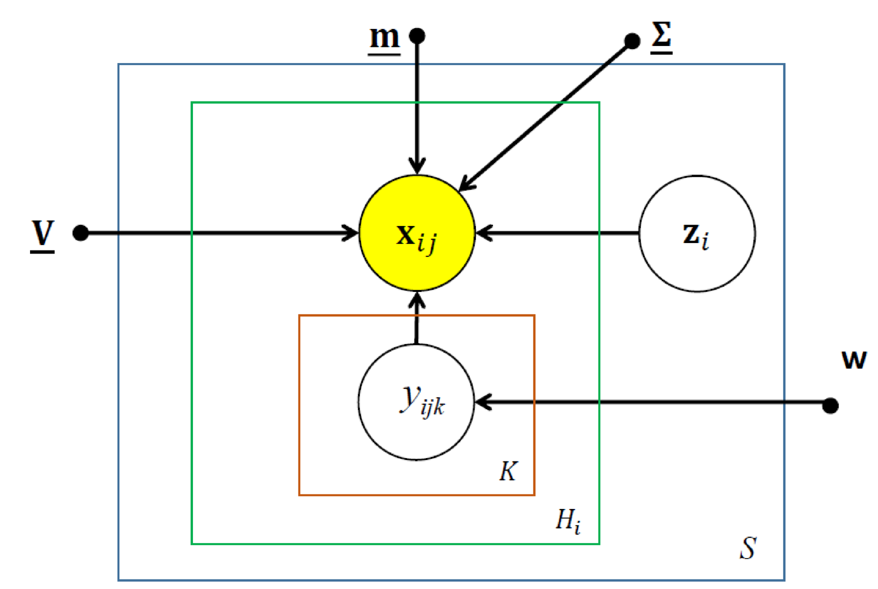

Fig.4: *Probabilistic graphical model of DNN-mPLDA.*

E-Step:

$$\langle y_{ijk} | \mathbf{x}_{ij} \rangle = \gamma_{\mathbf{x}_{ij}}(y_{ijk})$$

$$\mathbf{L}_i = \mathbf{I} + \sum_{k=1}^{K} N_{ik} \mathbf{V}_k^{\mathsf{T}} \mathbf{\Sigma}_k^{-1} \mathbf{V}_k$$

$$\langle \mathbf{z}_i | \mathcal{X} \rangle = \mathbf{L}_i^{-1} \sum_{k=1}^{K} \sum_{j=1}^{H_i} \langle y_{ijk} | \mathbf{x}_{ij} \rangle \mathbf{V}_k^{\mathsf{T}} \mathbf{\Sigma}_k^{-1} (\mathbf{x}_{ij} - \mathbf{m}_k)$$

$$\langle \mathbf{z}_i \mathbf{z}_i^{\mathsf{T}} | \mathcal{X} \rangle = \mathbf{L}_i^{-1} + \langle \mathbf{z}_i | \mathcal{X} \rangle \langle \mathbf{z}_i | \mathcal{X} \rangle^{\mathsf{T}}$$

M-Step:

$$\mathbf{m}'_k = \frac{\sum_{i=1}^{S} \sum_{j=1}^{H_i} \langle y_{ijk} | \mathbf{x}_{ij} \rangle \mathbf{x}_{ij}}{\sum_{i=1}^{S} \sum_{j=1}^{H_i} \langle y_{ijk} | \mathbf{x}_{ij} \rangle}$$

$$\mathbf{V}'_k = \left\{ \sum_{i=1}^{S} \sum_{j=1}^{H_i} \left[\langle y_{ijk} | \mathbf{x}_{ij} \rangle (\mathbf{x}_{ij} - \mathbf{m}'_k) \langle \mathbf{z}_i | \mathcal{X} \rangle^{\mathsf{T}} \right] \right\} \left[\sum_{i=1}^{S} N_{ik} \langle \mathbf{z}_i \mathbf{z}_i^{\mathsf{T}} | \mathcal{X} \rangle \right]^{-1}$$

$$\mathbf{\Sigma}'_k = \frac{1}{\sum_i N_{ik}} \sum_{i=1}^{S} \sum_{j=1}^{H_i} \Big[\langle y_{ijk} | \mathbf{x}_{ij} \rangle (\mathbf{x}_{ij} - \mathbf{m}'_k)(\mathbf{x}_{ij} - \mathbf{m}'_k)^{\mathsf{T}} - \mathbf{V}'_k \langle \mathbf{z}_i | \mathcal{X} \rangle \langle y_{ijk} | \mathbf{x}_{ij} \rangle (\mathbf{x}_{ij} - \mathbf{m}'_k)^{\mathsf{T}} \Big].$$

Likelihood Ratio Scores:

$$S_{\text{DNN-mPLDA}}(\mathbf{x}_s, \mathbf{x}_t) = \frac{\sum_{k_s=1}^{K} \sum_{k_t=1}^{K} \gamma_{\mathbf{x}_s}(y_{k_s}) \gamma_{\mathbf{x}_t}(y_{k_t}) \mathcal{N}\left([\mathbf{x}_s^{\mathsf{T}} \ \ \mathbf{x}_t^{\mathsf{T}}]^{\mathsf{T}} \mid [\mathbf{m}_{k_s}^{\mathsf{T}} \ \ \mathbf{m}_{k_t}^{\mathsf{T}}]^{\mathsf{T}}, \hat{\mathbf{V}}_{k_s k_t} \hat{\mathbf{V}}_{k_s k_t}^{\mathsf{T}} + \hat{\mathbf{\Sigma}}_{k_s k_t} \right)}{\left[\sum_{k_s=1}^{K} \gamma_{\mathbf{x}_s}(y_{k_s}) \mathcal{N}\left(\mathbf{x}_s | \mathbf{m}_{k_s}, \mathbf{V}_{k_s} \mathbf{V}_{k_s}^{\mathsf{T}} + \mathbf{\Sigma}_{k_s} \right) \right] \left[\sum_{k_t=1}^{K} \gamma_{\mathbf{x}_t}(y_{k_t}) \mathcal{N}\left(\mathbf{x}_t | \mathbf{m}_{k_t}, \mathbf{V}_{k_t} \mathbf{V}_{k_t}^{\mathsf{T}} + \mathbf{\Sigma}_{k_t} \right) \right]}$$

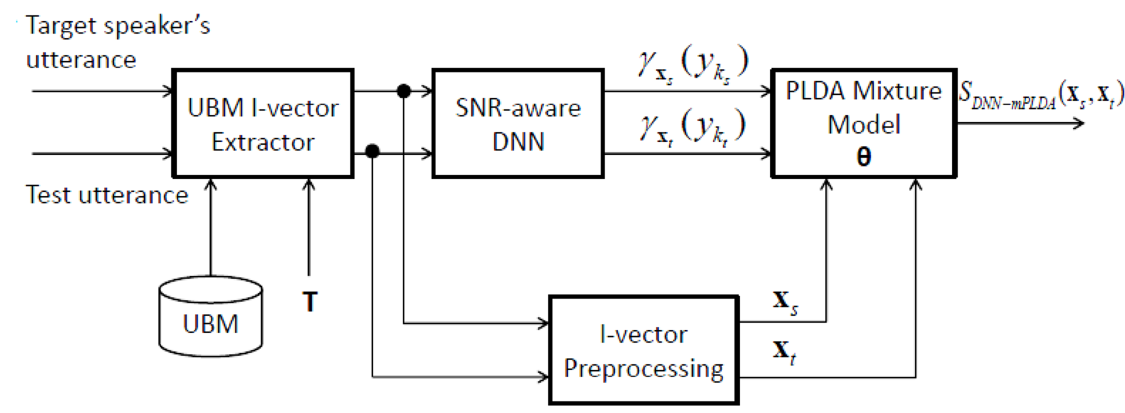

Fig.5: *The scoring process in DNN-mPLDA.*

Results

- **Table2:** *Performance of PLDA, SI-mPLDA, SD-mPLDA and Proposed DNN-mPLDA on CC4*

Method	K	Male		Female	
		EER(%)	minDCF	EER(%)	minDCF
PLDA	-	3.49	0.308	3.14	0.353
SI-mPLDA	2	3.49	0.303	3.11	0.350
	3	3.31	0.302	3.02	0.351
	4	3.31	0.299	3.00	0.354
SD-mPLDA	2	3.37	0.307	3.13	0.359
	3	3.06	0.315	**2.65**	**0.331**
	4	3.20	0.311	2.88	0.334
DNN-mPLDA	2	**2.95**	**0.296**	2.77	0.346
	3	3.03	0.305	2.77	0.339
	4	3.10	0.319	2.79	0.347

- **Table3:** *Performance of PLDA, SI-mPLDA, SD-mPLDA and Proposed DNN-mPLDA on CC5*

Method	K	Male		Female	
		EER(%)	minDCF	EER(%)	minDCF
PLDA	-	2.97	0.290	2.47	0.346
SI-mPLDA	2	3.04	0.300	2.55	0.340
	3	3.06	0.286	2.41	0.345
	4	2.93	0.288	2.60	0.332
SD-mPLDA	2	2.92	0.298	2.50	0.344
	3	2.80	**0.276**	2.38	**0.324**
	4	2.87	0.284	2.38	0.347
DNN-mPLDA	2	2.86	0.282	2.38	0.326
	3	**2.73**	0.279	**2.36**	0.333
	4	2.78	0.278	2.38	0.329

References:

•M. W. Mak, X. M. Pang, and J. T. Chien "mixture of PLDA for noise robust i-vector speaker verification," *IEEE/ACM Trans. on Audio, Speech and Language Processing*, vol. 24, no. 1, pp. 130–142, 2016.

•Y. Lei, N. Scheffer, L. Ferrer, and M. McLaren "A novel scheme for speaker recognition using a phonetically-aware deep neural network," in Proc. ICASSP, pp. 1695-1699, 2014.