Towards PLDA-RBM based Speaker Recognition in Mobile Environment:

Designing Stacked/Deep PLDA-RBM Systems

A. Nautsch[⋆], H. Hao[⋆†], T. Stafylakis[‡], C. Rathgeb[⋆], C. Busch[⋆]

[⋆]Hochschule Darmstadt, CASED, da/sec Security Research Group
[†]Technical University of Denmark
[‡]Centre de Recherche Informatique de Montréal (CRIM)

Shanghai, 23.03.2016

Outline

1. Introduction
2. Proposing stacked/deep designs for PLDA-RBM
3. Experimental results on mobile data
4. Conclusion

Motivation

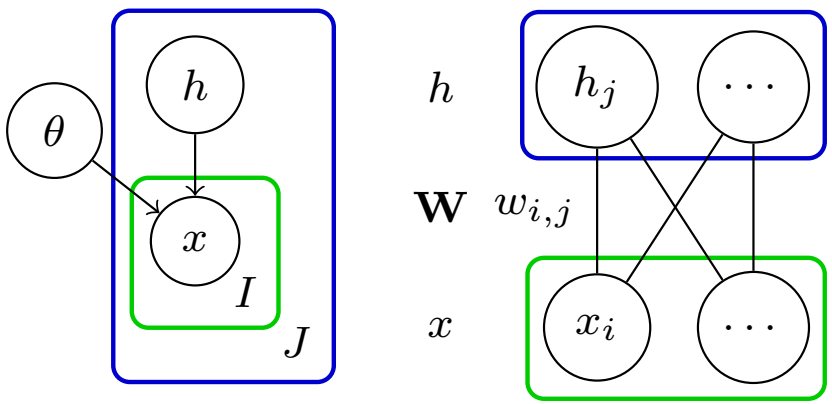
- MOBIO SRE'13 [Khoury+13]: limited mobile background data, i.e. state-of-the-art i-vector & PLDA performs insufficiently
- Restricted Boltzman Machines (RBMs) as PLDA-analogue with two-layer, undirected graphical models [Stafylakis+12]
- Deep Learning is more and more utilized in Speaker Recognition for robust estimation of feature spaces
- Exploiting i-vector reconstruction by deep PLDA-RBM design, i.e.: recovering biometric information

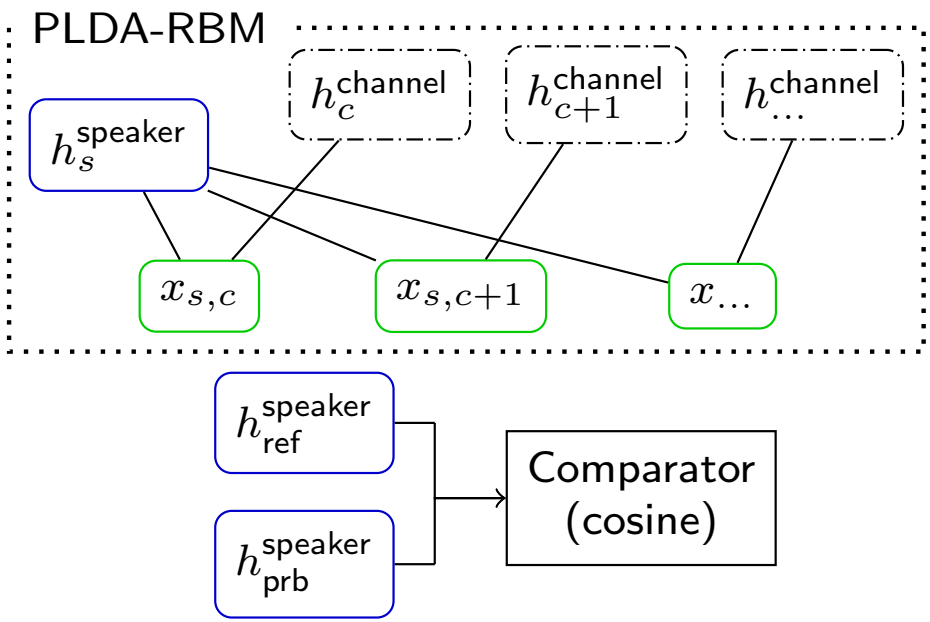
[Khoury+13] E. Khoury et al.: *The 2013 Speaker Recognition Evaluation in Mobile Environment*, IAPR ICB, 2013.

[Stafylakis+12] T. Stafylakis, P. Kenny, M. Senoussaoui, P. Dumouchel: *PLDA using Gaussian Restricted Boltzmann Machines with Application to Speaker Verification*, ISCA Interspeech, 2012.

Revisiting PLDA, RBMs and PLDA-RBM

- PLDA: decomposing i-vectors x into speaker h and channel θ factors
- RBM: bipartite undirected graphical model, w/o same-layer connections, with visible units x and hidden units h
- PLDA-RBM: decomposing speaker h_s^{speaker} and channel h_c^{channel} units





[Stafylakis+12] T. Stafylakis, P. Kenny, M. Senoussaoui, P. Dumouchel: *PLDA using Gaussian Restricted Boltzmann Machines with Application to Speaker Verification*, ISCA Interspeech, 2012.

Types and roles of Energy Functions

- Modeling the distribution of visible and hidden layers
- Relevant for updating weights $\boldsymbol{W}$
- Bernoulli-based RBMs
  - Visible layer: binary variables e.g., hand-writing digit data
  - Hidden layer: efficient classification features
- Gaussian-based RBMs
  - Visible layer: real-value, continuous data
  - Hidden layer: G-PLDA alike Gaussian sub-spaces, LLR scoring

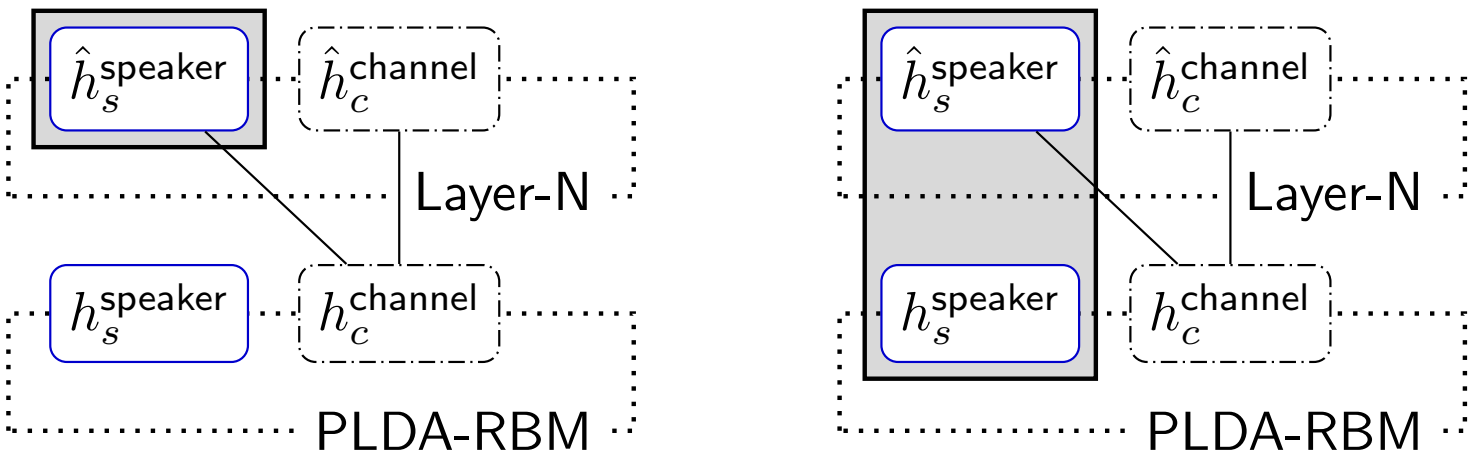
$\Rightarrow$ Gaussian-Gaussian (GG) and Gaussian-Bernoulli (GB) RBMs

[Stafylakis+12] T. Stafylakis, P. Kenny, M. Senoussaoui, P. Dumouchel: *PLDA using Gaussian Restricted Boltzmann Machines with Application to Speaker Verification*, ISCA Interspeech, 2012.

[Yamashita+14] T. Yamashita, M. Tanaka, E. Yoshida, Y. Yamauchi, H. Fujiyoshi: *To be Bernoulli or to be Gaussian, for a Restricted Boltzmann Machine*, IAPR IEEE ICPR, 2014.

Designing stacking concepts: Deep PLDA-RBM

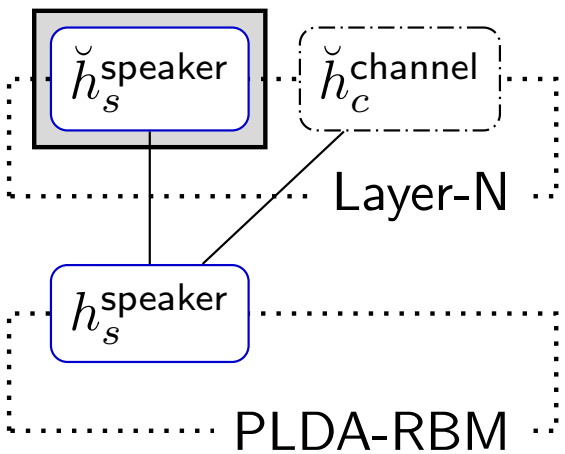
(a) Stacking on channel units



- Assumption: channel units still comprise biometric data
- Reconstructing biometric feature vectors
  - Selecting last reconstruction $\hat{h}_s^{\text{speaker}}$ of Layer-N
  - Concatenating all $\{h_s^{\text{speaker}}, \ldots, \hat{h}_s^{\text{speaker}}\}$

Designing stacking concepts: Deep PLDA-RBM

(b) Stacking on speaker units



- Assumption: speaker units still comprise non-biometric data
- Reconstructing biometric feature vectors:
  Selecting last reconstruction $\breve{h}_s^{\text{speaker}}$ of Layer-N

Experimental set-up

- PLDA-RBM
  - CD1 training, standard L2-regularization
  - Mini-batches of $\frac{1}{4}$ i-vectors/subject
  - Matlab implementation (MEDAL) [Stansbury13]
- Conducted analyses
  - Baseline comparison of G-PLDA [GarciaEspy11] to: GG PLDA-RBM [Stafylakis+12] and GB PLDA-RBM
  - Examining the impact of #units
  - Comparison of stacking concepts
- No calibration, performance reporting: C_{llr}^{min}

[Stansbury13] D.E. Stansbury: *Matlab Environment for Deep Architecture Learning (MEDAL) v0.1*, https://github.com/dustinstansbury/medal, 2013, [Online; accessed 2015-09-22].

[GarciaEspy11] D. Garcia-Romero, C. Y. Espy-Wilson: *Analysis of i-vector Length Normalization in Speaker Recognition Systems*, ISCA Interspeech, 2011.

[Stafylakis+12] T. Stafylakis, P. Kenny, M. Senoussaoui, P. Dumouchel: *PLDA using Gaussian Restricted Boltzmann Machines with Application to Speaker Verification*, ISCA Interspeech, 2012.

Experimental Results

Experimental set-up

- MOBIO Speaker Recognition Evaluation 2013 [Khoury+13]
- Database partitioning

Set	Female		Male	
	#Subjects	#Samples	#Subjects	#Samples
Background	13	2496	37	7104
dev-set (ref)	18	90	24	120
dev-set (prb)	18	1890	24	2520
eval-set (ref)	20	100	38	190
eval-set (prb)	20	2100	38	3990

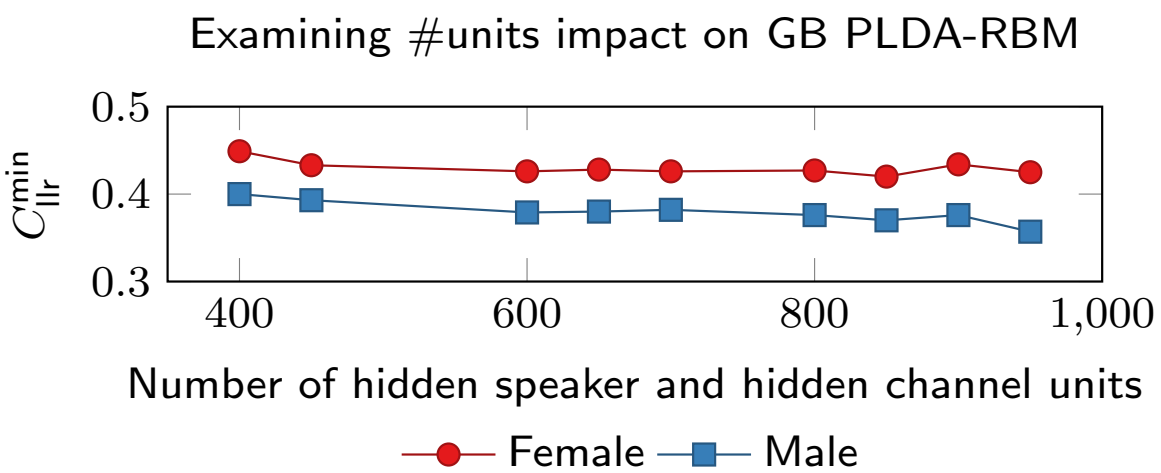
- Challenge: limited background data & mobile data,
  i.e.: LDA without significant benefits, thus no LDA
- MOBIO SRE'13 primary metric: HTER

[Khoury+13] E. Khoury et al.: *The 2013 Speaker Recognition Evaluation in Mobile Environment*, IAPR ICB, 2013.

Experimental Results

Baseline performance and #units

System	Female			Male		
	EER	FMR100	$C_{\text{llr}}^{\text{min}}$	EER	FMR100	$C_{\text{llr}}^{\text{min}}$
G-PLDA	15.3	63.5	0.488	**12.2**	**44.6**	**0.413**
PLDA-RBM — single layer						
GG	17.7	64.2	0.552	16.7	60.4	0.526
GB	**13.5**	**51.2**	**0.451**	12.3	48.3	0.418



Experimental Results

Comparison of stacking concepts

- Stacking concepts comparison w.r.t. #layers

#layers	Female		Male	
	(a) channel	(b) speaker	(a) channel	(b) speaker
1	0.420		0.370	
2	**0.392**	0.481	**0.341**	0.452
3	0.394	0.487	0.346	0.475

- Feature composition on (a) stacking channel units

	#layers	1	2	3	4	5
Female	single	0.420	0.505	0.551	0.691	0.715
	fused		**0.392**	0.394	0.398	0.394
Male	single	0.370	0.459	0.510	0.639	0.681
	fused		0.341	0.346	**0.340**	0.342

Experimental Results

Comparison to best systems of MOBIO SRE'13 having 1 sub-system

System	Female		Male	
	HTER	$C_{\text{llr}}^{\text{min}}$	HTER	$C_{\text{llr}}^{\text{min}}$
MOBIO-female (GMM – UBM)	11.6	n/a	9.1	n/a
MOBIO-male (GMM – UBM)	12.8	n/a	**8.9**	n/a
G-PLDA (gender-pooled)	16.4	0.522	9.9	0.326
1-layer GB PLDA-RBM	12.0	0.397	10.6	0.361
2-layer GB PLDA-RBM (channel-stacked)	**11.3**	**0.368**	9.0	**0.319**

Conclusion

- PLDA-RBM is applicable for (limited) mobile data
  - Benefits from GB assumption
  - GB PLDA-RBM outperforms conventional G-PLDA
- Recovering biometric information by exploiting deeper layers
  - Proposing stacking on channel units concept
  - Biometric information decreases in higher layers

  $\Rightarrow$ Accumulation of biometric data by $\{h_s^{\text{speaker}}, \ldots, \hat{h}_s^{\text{speaker}}\}$
- High computational efforts, providing more reliable evidence

- Perspectives
  - Examining large-scale NIST SRE databases
  - Robust RBM training e.g., drop-outs, fine-tuning
  - Deep layer designs e.g., GB-GB vs. GB-BB vs. GB-BG

This work has been funded by the Center for Advanced Security Research Darmstadt (CASED), and the Hesse government (project no. 467/15-09, BioMobile).

LOEWE
Exzellente Forschung für Hessens Zukunft

CASED

h_da
HOCHSCHULE DARMSTADT
UNIVERSITY OF APPLIED SCIENCES

Andreas Nautsch
Doctoral Researcher | Research Area: Secure Services

CASED
Mornewegstr. 32
64293 Darmstadt/Germany
andreas.nautsch@cased.de

Telefon +49 6151 16-75182
Fax +49 6151 16-4321
www.cased.de