PLDA using Gaussian Restricted Boltzmann Machines with application to Speaker Verification

T. Stafylakis, P. Kenny, M. Senoussaoui & P. Dumouchel

CRIM & ETS, Montreal(Quebec), Canada

September 12, 2012
Outline

- PLDA-based Speaker Verification
- Short introduction to Restricted Boltzmann Machines
- The proposed model for PLDA
- Experiments with i-vectors on NIST-2010
General about PLDA-based SV

- Let \( v \) be an i-vector (usually LDA projected and length-normalized)
- The standard (i.e. directed-graph) PLDA model
  \[
  v = Vs + Uc + \epsilon
  \]
  where \( s \sim N(0, I) \), \( c \sim N(0, I) \) and \( \epsilon \sim N(0, \Sigma_\epsilon) \)
- Scores based on LLR (some calibration usually required)
- A similar generative model can be assumed with undirected graphical models (at least this was our initial intension)
Restricted Boltzmann Machines (RBM)

An RBM...

- No connection between nodes of the same layer
- Allows fast training (blocked-Gibbs sampling)
- Correlations between nodes in $\mathbf{v}$ are still present in the marginal $P(\mathbf{v}|\mathbf{W})$
- The hidden variable $\mathbf{h}$ capture higher level information
Some useful expressions...

- The joint distribution (p.d.f. or p.m.f)

\[ \log P(v, h|\Theta) = \log P^*(v, h|\Theta) - \log Z(\Theta) \]

where

\[ \log P^*(v, h|\Theta) = -\sum_i \frac{(v_i - a_i)^2}{\sigma_i^2} - \sum_j \frac{(h_j - b_j)^2}{\sigma_j^2} + \sum_{i,j} v_i h_j W_{ij} \]

- Precision (i.e. inverse covariance)

\[ P = (I - W)\Sigma_{diag}^{-1/2} \]

where

\[ W = \begin{pmatrix} 0 & W \\ W^T & 0 \end{pmatrix} \]
Proposed PLDA model

For all (say $n_s$) recordings of a single speaker $s$

**Note:** $W_s = n_s^{-1/2} W_s$ and $\sigma_s = n_s^{-1/2}$ (assuming $\sigma_s = 1$).

**Figure:** (a) Introducing a compact illustration, (b) Proposed model
Learning with RBMs

Learning using Contrastive Divergence (CD)

- Split training set into speakers, **minibatches**, proceed with one speaker at a time, use several epochs
- **Contrastive Divergence** update formula:
  \[
  \Delta W \propto n_s \left( E_{P_{\text{data}}} [v_s h_s^T] - E_{P_T} [v_s h_s^T] \right)
  \]
- \( P_T \) a distribution defined by running a **Gibbs** chain, initialized at the data, for \( T \) full steps
Assume two i-vectors $v_i, i = 1, 2$, and let $s_i = W_s^T v_i$.

The log-likelihood ratio (LLR) between $H_1$ and $H_0$ is

$$LLR = \frac{1}{2} (s_1 - s_2)^T (s_1 - s_2) + cnt$$

By forcing **speaker factors** to lie on the surface of the sphere (i.e. length norm), the model is similar to the **cosine distance** scoring.

WCCN may also be applied.
Experiments

MFCC, UBM and i-vector configuration

- 20 dim. Gaussianized MFCC $+\Delta + \Delta\Delta$
- 2048-component, gender independent UBM (NIST ’04 & ’05 for enrollment data)
- 800-dim i-vectors (gender independent)
- Best performance of the proposed model with 80-dim speaker factors (compare that to 200-dim of standard LDA)
Results on NIST ’10 female (core-extended): PLDA vs. cosine distance
Experiments

Table: Results on NIST-10 female tel. data (core extended).

<table>
<thead>
<tr>
<th></th>
<th>EER (%)</th>
<th>MDCF-2008</th>
<th>MDCF-2010</th>
</tr>
</thead>
<tbody>
<tr>
<td>cosine distance</td>
<td>3.45</td>
<td>0.33</td>
<td>0.49</td>
</tr>
<tr>
<td>twin model⁠¹</td>
<td>2.51</td>
<td>0.27</td>
<td>0.41</td>
</tr>
<tr>
<td>Proposed (a)</td>
<td>2.75</td>
<td>0.28</td>
<td>0.46</td>
</tr>
<tr>
<td>Proposed (b)</td>
<td>2.85</td>
<td>0.27</td>
<td>0.43</td>
</tr>
</tbody>
</table>

Conclusions and challenges

Conclusions...

- An undirected graphical model for PLDA was presented, based on RBMs
- Can be seen either as a complete back-end system (like standard PLDA) or as dimensionality reduction model (like LDA)
- Performance comparable to standard PLDA and superior to LDA

Challenges...

- Incorporate the uncertainty of the i-vector estimates
- Use it as an alternative i-vector extractor...