An Unsupervised Dynamic Bayesian Network Approach to Measuring Speech Style Accommodation
M. Jain, et al. (EACL ‘12)

Keith Maki, F’16 11-776: HCMC
LTI, Carnegie Mellon University
Speech Style Accommodation

- Encompassed by Social Psychology of Language
- Conversational partners shift their speech style
  - Also referred to as entrainment, priming, adaptation
Approach

- Unsupervised model of accommodation
- Explicitly represent “accommodation state” as a latent variable ("relational gestalt")
  - Represents the influence partners have over each other’s language as a reflection of social processes
- Model style shift over time
Accommodation is reflected in prosody: pitch, energy, speaking rate, etc.

Analyze each speaker turn separately, with 50ms sliding window (40ms overlap)

Use OpenSMILE to extract 7 features:
- voice probability, harmonic to noise ratio,
  voice quality, F0, F raw 0, F env 0, and loudness

Normalize 10-bin histogram: 70-dim $o^s_t$
Dynamic Bayes Net

Figure 1: An example Dynamic Bayesian Network (DBN) showing the temporal relationship between three random variables \((A, B\) and \(C)\). \(A\) is observered and dependent on two hidden variables \(B\) and \(C\). Directed edges across time \((t - 1 \rightarrow t)\) indicate temporal relationships between variables. In this example, the variables \(A_t\) and \(B_t\) are both dependent on \(B_{t-1}\) with the relationship defined through conditional distributions \(P(A_t|B_{t-1})\) and \(P(B_t|B_{t-1})\).

Figure 2: The basic generative model.
Accommodation models

Figure 3: ISM: The dynamics of each speaker are independent of the other speaker.

Figure 4: CSDM: A speaker’s style depends on their partner’s style at the previous turn.

Figure 5: SASM: Both partners’ styles depend on mutual accommodation to one another.

Figure 6: AASM: Accommodation state associated with every speaker turn.

Figure 7: SASDM: A speaker’s style depends both on mutual accommodation and the partner’s style in the previous turn.

Figure 8: AASDM: The accommodation state associated with every speaker and a speaker’s style depends on the partner’s style.
Experiments

- Compare between real vs constructed dialogs (balanced)
  - Constructed had 1 real and 1 synthetic partner
  - Find support for claim that approaches modeling accommodation as process over time can better distinguish real vs constructed interactions

<table>
<thead>
<tr>
<th>Model</th>
<th>DI</th>
<th>II</th>
<th>Real $\mu(\sigma)$</th>
<th>Constructed $\mu(\sigma)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>SASDM</td>
<td>T</td>
<td>S</td>
<td>.54 (.23)</td>
<td>.44 (.29)</td>
</tr>
<tr>
<td>SASM</td>
<td>F</td>
<td>S</td>
<td>.54 (.23)</td>
<td>.44 (.29)</td>
</tr>
<tr>
<td>CSDM</td>
<td>T</td>
<td>F</td>
<td>.6 (.26)</td>
<td>.52 (.3)</td>
</tr>
<tr>
<td>ISM</td>
<td>F</td>
<td>F</td>
<td>.56 (.25)</td>
<td>.51 (.32)</td>
</tr>
<tr>
<td>AASM</td>
<td>F</td>
<td>A</td>
<td>.6 (.24)</td>
<td>.51 (.3)</td>
</tr>
<tr>
<td>AASDM</td>
<td>T</td>
<td>A</td>
<td>.61 (.24)</td>
<td>.48 (.3)</td>
</tr>
</tbody>
</table>

Table 1: Accommodation measured using different models. Legend: $\mu$=mean, $\sigma$ = standard deviation, DI = “Direct Influence”, II = “Indirect Influence”.
Experiments

- Across models, able to separate classes \((p < .0001)\)
- More accommodation in real pairs
- SASDM is the only model which individually separates the classes
  - Indirect links more important than direct links

http://dl.acm.org/citation.cfm?id=2380912