IDENTIFICATION OF TRANSACTIVITY IN GROUP CONVERSATION

- Transactivity: When a speaker B builds upon a statement of speaker A in a reasonable way to add knowledge about doing a particular task during the conversation.

- Our Problem: To identify “transactivity” using acoustic properties of speech.
WHY THE PROJECT MIGHT BE IMPORTANT

- Autonomous Tutoring
  - Computer rates whether a conversation was useful
  - Compares contribution of each speaker
  - Leads to → Computer assisted group tutoring

- Exploring the existence of a link between semantics and speech features.

- As a pre-processing tool for group conversation summarizers.

- Follow group chats and extract people’s biases → useful for advertising !!
**Experimental Setup**

- 3 people in conversation
  - Trying to build a machine to prevent egg from breaking when dropped from 3rd floor.

- Each Speaker ➔ Own Microphone

- 8 such meetings

- Speech ➔ Manually transcribed into text

- Annotations ➔ From prior research
FIRST ISSUE: EXTRACT NOISELESS SPEECH OF EACH SPEAKER

- Autonomous Speech Segmentation
  - Removing Cross-Talk
  - Removing Noise
  - Performance:
    - Fraction Missed ~ 2% (Except in speaker 2 recordings)
    - Extra Fraction Recorded ~ 30%

- Possible Solutions
  - Remove breathing using spectrum analysis: Hypothesis
  - Build a classifier based on Mfcc’s, Pitch etc.

Later ➔ We left the issue and trusted the time stamps in annotation. (although a lot of time was spent making others understand this)
HYPOTHESIS FOR THE OVERALL SYSTEM

- Extract acoustic features from speech recordings.
- Speech $\rightarrow$ Text
- Extract Semantic frames
- Acoustic Feature + Semantics $\rightarrow$ Transactivity
- Offline
- Want High Precision & Recall: Reflected in F-Score
**Initial Experiment**

- Involved only acoustic features:
  - Mfcc
  - Pitch
  - Energy
  - Duration
  - Speaker ID

- Svm-Hmm model → Reasoning vs NonReasoning

- Only 2 meetings
## Results

<table>
<thead>
<tr>
<th>exp#</th>
<th>training set</th>
<th>test set</th>
<th>features used</th>
<th>Recall (%)</th>
<th>Precision (%)</th>
<th>F-score (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>100% of meeting1</td>
<td>100% of meeting2</td>
<td>id, duration</td>
<td>40.68</td>
<td>43.64</td>
<td>42.11</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>id, duration, 1 pca</td>
<td>38.99</td>
<td>46.93</td>
<td>42.59</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>id, duration, 2 pca</td>
<td>49.15</td>
<td>46.03</td>
<td>47.54</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>id, duration, 3 pca</td>
<td>47.46</td>
<td>45.90</td>
<td>46.67</td>
</tr>
<tr>
<td>2</td>
<td>80% of (meeting1 + meeting2)</td>
<td>remaining 20% (meeting1 + meeting2)</td>
<td>id, duration</td>
<td>35.29</td>
<td>92.31</td>
<td>51.06</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>id, duration, 1 pca</td>
<td>44.12</td>
<td>83.33</td>
<td>57.69</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>id, duration, 2 pca</td>
<td>52.94</td>
<td>75</td>
<td>62.07</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>id, duration, 3 pca</td>
<td>52.94</td>
<td>75</td>
<td>62.07</td>
</tr>
</tbody>
</table>
Next... 

- Hoping results would improve/generalize as we increase meetings

- Data priors: 17.36% relevant sentences
  - M2: 34.75%
  - M3: 8.18%
  - M4: 21.86%
  - M5: 11.31%
  - M6: 20%
  - M7: 9.59%
  - M8: 16.87%
  - M9: 20.24%
# Results SVM-HMM Model

Features: First 3 PCA, SpkId, Length

<table>
<thead>
<tr>
<th>Test on (Meeting)</th>
<th>Recall</th>
<th>Precision</th>
<th>F-Score</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M2</td>
<td>0.107</td>
<td>0.92</td>
<td>0.193</td>
</tr>
<tr>
<td>M3</td>
<td>0.10</td>
<td>0.67</td>
<td>0.176</td>
</tr>
<tr>
<td>M4</td>
<td>0.365</td>
<td>0.652</td>
<td>0.469</td>
</tr>
<tr>
<td>M5</td>
<td>0.272</td>
<td>0.442</td>
<td>0.337</td>
</tr>
<tr>
<td>M6</td>
<td>0.161</td>
<td>0.678</td>
<td>0.260</td>
</tr>
<tr>
<td>M7</td>
<td>0.111</td>
<td>0.381</td>
<td>0.172</td>
</tr>
<tr>
<td>M8</td>
<td>0.178</td>
<td>0.667</td>
<td>0.322</td>
</tr>
<tr>
<td>M9</td>
<td>0.148</td>
<td>0.937</td>
<td>0.256</td>
</tr>
</tbody>
</table>
CHANGE OF MODEL TO SVM

- Features: 3 PCA. SpkId, Duration
- Various kernels tried: almost same performance

<table>
<thead>
<tr>
<th>Test on (Meeting)</th>
<th>Recall</th>
<th>Precision</th>
<th>F-Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>M2</td>
<td>0.122</td>
<td>0.896</td>
<td>0.214</td>
</tr>
<tr>
<td>M3</td>
<td>0.203</td>
<td>0.60</td>
<td>0.304</td>
</tr>
<tr>
<td>M4</td>
<td>0.33</td>
<td>0.63</td>
<td>0.43</td>
</tr>
<tr>
<td>M5</td>
<td>0.19</td>
<td>0.46</td>
<td>0.27</td>
</tr>
<tr>
<td>M6</td>
<td>0.17</td>
<td>0.61</td>
<td>0.264</td>
</tr>
<tr>
<td>M7</td>
<td>0.088</td>
<td>0.29</td>
<td>0.135</td>
</tr>
<tr>
<td>M8</td>
<td>0.20</td>
<td>0.52</td>
<td>0.294</td>
</tr>
<tr>
<td>M9</td>
<td>0.06</td>
<td>1.00</td>
<td>0.112</td>
</tr>
</tbody>
</table>
IN A NUTSHELL...

- Train on M2,M9,M6: Better Results.

- Train on M3,M7: very poor
  - Almost 0 recall

- Jack-Knife
  - Not very encouraging

- Jack-Knife: Using Only SVM
  - Linear, Polynomial (Order 3), and radial basis: Kernel
  - Biasing didn’t help much either
  - Very Slight improvement
TRYING FOR SEMANTIC FEATURES

- Frame Semantic Approach

- What does it do?
  - Tries to identify context
  - Based on identifying key words and context dependent attributes

- Extract Frames \(\rightarrow\) Capture Concepts/Context

- Sequence of Concepts \(\rightarrow\) Capture Transactivity

- Never worked out 😞
Ada Boost: The Simple & Elegant classifier !! 😊

- Features
  - MFCC
  - Energy
  - Amplitude
  - Duration of Segment
  - Speaker ID
  - Phoneme Histogram
  - Phoneme Rate
  - Phoneme Count
  - Pitch
  - Stylistic Features
    - KL distance between phoneme histograms
  - Language Model Probabilities
**Twist in the Tale..**

- Change in Gold standard time stamps !

- Bigger Twist: Change in Labels

- 3 classes
  - Transactivity
  - Externalization
  - Neither

- Ada Boost was used to make 3 binary classifiers
EXPERIMENT

- The bias of boosting classifier was varied from -1 to 1
- Looking for EER, Kappa and F-Score
- Some improvement in results!!
REASONING VS NON REASONING

- Results
  - Bias: -0.4694
  - Kappa: 0.467
  - F-Score: 0.561
  - Recall: 0.631
  - Precision: 0.51

- EER = 0.22

- Best Features
  - Length
  - Phoneme Rate
  - Some Acoustic features (Energy, MFCC)
FEATURE PERFORMANCE
EER Graph
**Transactivity Vs Rest**

- **Results**
  - Bias: -1.157
  - F-score: 0.357
  - Kappa: 0.2301
  - Recall: 0.72

- **EER ~ 0.21**

- **Best Features**
  - Duration
  - Phoneme ‘B’
  - Acoustics: Mfcc, Energy
FEATURE PERFORMANCE
EER Graph
EXTERNALIZATION VS REST

Results

• Bias: -1.15
  - Kappa: 0.255
  - F-Score: 0.328
  - Recall: 0.702

• Bias: -0.84
  - Kappa: 0.259
  - F-Score: 0.323
  - Recall: 0.4785

EER ~ 0.22

Best Features

• Duration
• Other Features: Little contribution
FEATURE PERFORMANCE
EER Graph

![EER Graph Image]

- X: 0.22
- Y: 0.78
**IDEAS...**

- Evolution of transactivity over time
  - HMM Models

- Use of stylistic features
  - Variation in tone, loudness, choice of words
    - As compared to other speakers
    - Compared to current speaker’s previous utterances
    - Normalize these features for each speaker.

- Combining Semantics in some way
  - Frame semantic sounds promising.

- Use of syllables instead of phonemes.

- Classifier between externalization v/s Transactivity.
Questions ??

Comments ??  Suggestions ??
THANK YOU !!