Agreement Detection in Multiparty Conversations

FEAST,
21st October 2009

Sebastian Germesin
Theresa Wilson
Motivation

- Growing interest in extracting and summarising information from meetings
- One important type of information are agreements / disagreements

⇒ Development of an automatic detection system!
Agreements in this work

- What do we mean by “(dis)agreements”?  
- Utterances where a speaker agrees/disagrees with an idea/opinion/sentiment of another speaker [Wilson, 2008]
- Agreements in the context of multi-party conversations
Example:

...  
\textit{Industrial Designer:} Finding them is really a pain.  
\textit{Marketing Expert:} Hm.  
\textit{Industrial Designer:} I mean, when you want it, it’s kicked under the table or so.  
\textit{Project Manager:} Yeah, that’s right.  
...
Data (AMI Corpus)

AMI meeting corpus

- 100 hours of audio and video recorded meetings
- 4 participants
- (guided) task: “Design a remote control!”
- Variety of annotations, e.g.:
  - Transcribed speech, ASR output, ...
  - Dialogue Acts, Disfluencies, ...
  - Head- & Hand-Gestures, VFOA, ...
  - and...
(Dis)Agreement Annotations

- 20 AMI meetings have been annotated
- Word-based annotation scheme
- 16 for training, 4 for evaluation
- 636 agreements / 70 disagreements
- Aligning to DA segments to preserve comparability with ICSI research ([Hillard 03], [Galley 04], [Hahn 06]):
  - 19,043 segments
  - 876 segments contain agreements
  - 118 segments contain disagreements
  \[ 4.6\% : 0.6\% : 94.8\% \text{ (agree : disagree : other)} \]
Automatic Detection System

“A Finding them is really a pain!”

“Yeah”

Agree Detection

“agreement”

Target Speaker Detection

B agrees with A
Automatic Detection System

Agreement Detection

Use HPR-output as feature

High-Precision Rules

Decision Tree

Cond. Random Field

“agreement”, “disagreement”, “other”

Cope with skewed classes

German Research Center for Artificial Intelligence GmbH
Automatic Detection System

HPRs - High Precision Rules
- Implement prior knowledge in simple set of rules before actual classification
- Reduce data skewness
- Rule-Types:
  - Target-Content
  - Dialogue Act (DA) Labels
  - Subjective Content
  - N-grams
Automatic Detection System

Decision Tree
- C4.5 implementation from WEKA Toolkit
- Lexical features
- Prosodic features
- Structural features
- HPR-output
- Contextual features

Conditional Random Fields
- CRF implementation from Stanford NER
- Lexical features
- Prosodic features
- Structural features
- HPR-output
### Experimental Results

#### High-Precision-Rules

<table>
<thead>
<tr>
<th>name</th>
<th>correct</th>
<th>wrong</th>
</tr>
</thead>
<tbody>
<tr>
<td>No-Target</td>
<td>740</td>
<td>12</td>
</tr>
<tr>
<td>DA-Label (src)</td>
<td>295</td>
<td>2</td>
</tr>
<tr>
<td>DA-Label (tar)</td>
<td>274</td>
<td>2</td>
</tr>
<tr>
<td>Silence</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Length</td>
<td>141</td>
<td>5</td>
</tr>
<tr>
<td>Pre-Class.</td>
<td>1890</td>
<td>0</td>
</tr>
<tr>
<td>Agreement</td>
<td>4</td>
<td>0</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>agree</th>
<th>other</th>
<th>“unclass”</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>3362</td>
<td>554</td>
</tr>
</tbody>
</table>

3920 segments
## Experimental Results

### Agreement Detection

<table>
<thead>
<tr>
<th></th>
<th>Baseline</th>
<th>CRFs</th>
<th>DTs</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>w/ HPRs</td>
<td>w/o HPRs</td>
<td>w/ HPRs</td>
</tr>
<tr>
<td>Acc [%]</td>
<td>97.8</td>
<td>98.0</td>
<td>98.1</td>
</tr>
<tr>
<td>Prec [%]</td>
<td>0.0</td>
<td>57.6</td>
<td>58.8</td>
</tr>
<tr>
<td>Rec [%]</td>
<td>0.0</td>
<td>36.3</td>
<td>34.6</td>
</tr>
<tr>
<td>F1 [%]</td>
<td>0.0</td>
<td>44.5</td>
<td>43.5</td>
</tr>
<tr>
<td>Kappa</td>
<td>0.0</td>
<td>0.40</td>
<td>0.39</td>
</tr>
<tr>
<td>RT Factor</td>
<td>0.0</td>
<td>0.005</td>
<td>0.005</td>
</tr>
</tbody>
</table>
Automatic Detection System

“A Finding them is really a pain!”

“Yeah”

Target Speaker Detection

B agrees with A

“agreement”

A

B

Agree Detection
Automatic Detection System

“Finding them is really a pain!”

“Yeah”

Agree Detection

“agreement”

Target Speaker Detection

B agrees with A

A

B
Target Speaker Detection:

- Novelty in agreement detection
- Preliminary experiments using Adjacency Pair-Annotation
Internal Representation:

- Use speaker-dependent (relative) labels
  - 0 for current speaker
  - 1 for previous speaker
  - ...

- *Let’s see this in the example:*
Example:

Example:

Index ‘1’:
Finding them is really a pain.

Index ‘2’:
Hm.

Index ‘1’:
I mean, when you want it, it’s kicked under the table or so.

Index ‘0’:
Yeah, that’s right.

agreement

addressee

target
Automatic Detection System

![Bar chart showing the percentage of agreement and disagreement for speaker indices 1, 2, and 3.]

- **Speaker Index 1**: Agree 70%, Disagree 30%
- **Speaker Index 2**: Agree 50%, Disagree 50%
- **Speaker Index 3**: Agree 10%, Disagree 90%
Automatic Detection System

Target Detection:
- Use structural information from Adjacency Pairs to improve target speaker detection! (backward-window of 10 segments)
- Fall back to speaker ‘1’ if no AP is available
### Experimental Results

#### Baseline

<table>
<thead>
<tr>
<th>classified as</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>Acc</th>
<th>F₁</th>
<th>k</th>
</tr>
</thead>
<tbody>
<tr>
<td>real</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>164</td>
<td>0</td>
<td>0</td>
<td>78.0</td>
<td>64.5</td>
<td>0.00</td>
</tr>
<tr>
<td>2</td>
<td>78</td>
<td>0</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>12</td>
<td>0</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
## Experimental Results

### Using AP-Information

<table>
<thead>
<tr>
<th>classified as</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>Acc</th>
<th>F_1</th>
<th>k</th>
</tr>
</thead>
<tbody>
<tr>
<td>real</td>
<td>1</td>
<td>163</td>
<td>0</td>
<td>1</td>
<td>86.9</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>38</td>
<td>40</td>
<td>0</td>
<td>67.2</td>
<td>0.52</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>10</td>
<td>1</td>
<td>1</td>
<td>14.2</td>
<td></td>
</tr>
</tbody>
</table>

- 56% improvement
Conclusion

Developed a system for agreement detection:

- Utilized a variety multi-modal, heterogeneous features (e.g., lexical, prosodic, structural)
- Investigated the use of High-Precision Rules to deal with imbalanced class distribution
- Evaluated two different types of machine learning techniques
  - Conditional Random Fields
  - Decision Trees
- Accuracy: 98.1%
- Kappa: 0.40
- CRF: higher Precision
- DT: higher Recall
Novelty: Target Speaker detection!

- Introduced preliminary approach, using structural information from the adjacency pairs
- 56% relative improvement over the baseline
- Kappa value of 0.52
Outlook

- Separate detection of agreements and disagreements
- Separate detection of one-word and multi-word agreements
- Use machine learning for addressee detection
- Use automatic annotations
- Use other features (e.g., visual cues)
- Care about data skewness
Thank you!