Opinions in Interactions: New Annotations of the SEMAINE Database
Valentin Barriere¹, Slim Essid², Chloé Clavel²
¹European Commission’s Joint Research Center - Ispra, ²Télécom Paris

Overview

- We present the process to collect new annotations of opinion over the multimodal corpus SEMAINE database composed of dyadic interactions [1]
- Using interactional context by seeing and annotation the whole conversation
- Using multimodal context by reading the text and listening to the audio recording at the same time
- We propose a baseline for the detection of opinions in interactions, reaching a F1 of 0.72.

SEMAINE Database

- 79 sessions, composed of 5,627 speech turns and 74k words, corresponding to 6h20 of dyadic interactions
- Between an user and an operator playing the role of a virtual agent
- Emotionally colored conversations annotated in a continuous way in Valence, Arousal, Power and Expectancy

Annotation Platform: integrate interactional/multimodal context

- Homemade php annotation platform
- Dialogic context: Conversational history and a contextual Adjacency Pair (pair of speech turns)
- Multimodal context: Audio and Text were aligned using [3] in order to use both the modalities to annotate

Annotation Platform: minimize the cognitive load

- Preliminary phase to annotate a dummy conversation
- Incrementally complex questions in order to reduce the cognitive load
- Special attention on mixed opinions cases to increase inter-annotator agreement opinions

Contact Information

- Email: valentin.barriere@ec.europa.eu
- Phone: +39 033278-3859

Annotations – Inter Annotator Agreement & Aggregation

- Given the low inter-annotator agreement using 4 classes, unreliable to use Mixed label as a fourth class.
- Using the prominent opinion allows to reach higher IAA.

Table 1: Krippendorff’s α per discussion using the Prominent label

<table>
<thead>
<tr>
<th>Speaker</th>
<th>μ (σ²)</th>
<th>α 3 classes</th>
<th>min</th>
<th>max</th>
<th>med</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Agent</td>
<td>80.4 (16.1)</td>
<td>54.9 (15.0)</td>
<td>85.9</td>
<td>56.0</td>
<td>66.3</td>
<td></td>
</tr>
<tr>
<td>User</td>
<td>46.7 (3.2)</td>
<td>60.9 (11.0)</td>
<td>39.1</td>
<td>58.1</td>
<td>66.3</td>
<td></td>
</tr>
<tr>
<td>All</td>
<td>39.0 (5.2)</td>
<td>60.5 (16.1)</td>
<td>100</td>
<td>80.5</td>
<td>66.3</td>
<td></td>
</tr>
</tbody>
</table>

Table 2: Opinion per speech turn obtained after aggregation using majority vote

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Valence</td>
<td>-3.6</td>
<td>18.7</td>
<td>4.3</td>
<td>11.9 (5.9)</td>
</tr>
<tr>
<td>Arousal</td>
<td>-11.5</td>
<td>-6.1</td>
<td>5.8</td>
<td>-7.9 (5.5)</td>
</tr>
<tr>
<td>Dominance</td>
<td>39.0</td>
<td>46.7</td>
<td>3.2</td>
<td>39.0 (5.2)</td>
</tr>
<tr>
<td>Surprise</td>
<td>33.8</td>
<td>31.4</td>
<td>0.9</td>
<td>32.7 (1.1)</td>
</tr>
</tbody>
</table>

Table 3: Statistics between the emotion-related and the opinion annotation

Correlations between labels

- There is a correlation between Valence and Opinion.
- Positive opinions are more likely to have high valence and high arousal
- Negative opinions are more likely to have low valence and low arousal
- Positive opinions are more dominant than neutral and negative ones.

Baseline

- We provide a multimodal baseline, enhancing SEMAINE’s state-of-the-art DialogueRNN model [4] using:
  - RoBERTa [5] speech turn embeddings as textual features
  - Comparé feature set [6] as audio feature
- Given the low inter-annotator agreement using 4 classes, allows to reach higher IAA.

Table 4: Baseline results on different tasks

<table>
<thead>
<tr>
<th>Method</th>
<th>Valence</th>
<th>Arousal</th>
<th>Opinion</th>
</tr>
</thead>
<tbody>
<tr>
<td>DialogueRNN</td>
<td>0.171</td>
<td>0.37</td>
<td>0.164</td>
</tr>
<tr>
<td>Our model</td>
<td>0.132</td>
<td>0.76</td>
<td>0.135</td>
</tr>
</tbody>
</table>

Table 4: Baseline results on different tasks

Conclusion & Future Work

- Collected opinion annotations per speech turns on SEMAINE
- Rich in opinions: 48.08% of the speech turns
- Can be used with continuous emotional annotations of the AVEC-2012 challenge of [6]

References

- Comparé feature set [6] as audio feature
- Given the low inter-annotator agreement using 4 classes, allows to reach higher IAA.

Table 4: Baseline results on different tasks

<table>
<thead>
<tr>
<th>Method</th>
<th>Valence</th>
<th>Arousal</th>
<th>Opinion</th>
</tr>
</thead>
<tbody>
<tr>
<td>DialogueRNN</td>
<td>0.171</td>
<td>0.37</td>
<td>0.164</td>
</tr>
<tr>
<td>Our model</td>
<td>0.132</td>
<td>0.76</td>
<td>0.135</td>
</tr>
</tbody>
</table>

Table 3: Statistics between the emotion-related and the opinion annotation

Correlations between labels

- There is a correlation between Valence and Opinion.
- Positive opinions are more likely to have high valence and high arousal
- Negative opinions are more likely to have low valence and low arousal
- Positive opinions are more dominant than neutral and negative ones.

Baseline

- We provide a multimodal baseline, enhancing SEMAINE’s state-of-the-art DialogueRNN model [4] using:
  - RoBERTa [5] speech turn embeddings as textual features
  - Comparé feature set [6] as audio feature
- Given the low inter-annotator agreement using 4 classes, allows to reach higher IAA.

Table 4: Baseline results on different tasks

<table>
<thead>
<tr>
<th>Method</th>
<th>Valence</th>
<th>Arousal</th>
<th>Opinion</th>
</tr>
</thead>
<tbody>
<tr>
<td>DialogueRNN</td>
<td>0.171</td>
<td>0.37</td>
<td>0.164</td>
</tr>
<tr>
<td>Our model</td>
<td>0.132</td>
<td>0.76</td>
<td>0.135</td>
</tr>
</tbody>
</table>

Table 4: Baseline results on different tasks

Conclusion & Future Work

- Collected opinion annotations per speech turns on SEMAINE
- Rich in opinions: 48.08% of the speech turns
- Can be used with continuous emotional annotations of the AVEC-2012 challenge of [6]

References

- Comparé feature set [6] as audio feature
- Given the low inter-annotator agreement using 4 classes, allows to reach higher IAA.

Table 4: Baseline results on different tasks

<table>
<thead>
<tr>
<th>Method</th>
<th>Valence</th>
<th>Arousal</th>
<th>Opinion</th>
</tr>
</thead>
<tbody>
<tr>
<td>DialogueRNN</td>
<td>0.171</td>
<td>0.37</td>
<td>0.164</td>
</tr>
<tr>
<td>Our model</td>
<td>0.132</td>
<td>0.76</td>
<td>0.135</td>
</tr>
</tbody>
</table>