Inducing Positive Perspectives with Text Reframing

Caleb Ziems, Minzhi Li, Anthony Zhang, Diyi Yang
Georgia Institute of Technology, National University of Singapore
cziems@gatech.edu

1. Introduction
- We introduce a new NLP task: **Positive Reframing**
- **Goal:** increase positivity without contradicting the meaning of the original utterance (premise)
- Unlike Sentiment Style Transfer, **Positive Reframing** must conversationally implicate the premise

2. Theoretical Framework: Positive Psychology
**Idea:** one can **overcome adversity** by shifting and rephrasing **cognitive distortions**
- Growth Mindset
- Impermanence
- Neutralizing
- Optimism
- Self-Affirmation
- Thankfulness

3. Data Annotation
- 1: Use distant supervision to identify negative posts
- 2: Apply Content Filters
- 3: Crowdworkers Annotate Data

4. Positive Psychology Frames Corpus

<table>
<thead>
<tr>
<th>ICC</th>
<th>Strat</th>
<th>Distribution</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.59</td>
<td></td>
<td>25.4%</td>
</tr>
<tr>
<td>0.60</td>
<td></td>
<td>19.5%</td>
</tr>
<tr>
<td>0.32</td>
<td></td>
<td>36.1%</td>
</tr>
<tr>
<td>0.44</td>
<td></td>
<td>48.7%</td>
</tr>
<tr>
<td>0.42</td>
<td></td>
<td>10.1%</td>
</tr>
<tr>
<td>0.42</td>
<td></td>
<td>13.0%</td>
</tr>
</tbody>
</table>

5. Automatic Reframing
- **Few-Shot** is positive but not as meaning-preserving
- **Supervised** can preserve meaning but not perfectly
  - Best performance when we condition the generation on the gold-standard strategies (Control)
  - We can also Predict the appropriate strategies

<table>
<thead>
<tr>
<th>Model</th>
<th>Meaning</th>
<th>Positivity</th>
<th>Fluency</th>
</tr>
</thead>
<tbody>
<tr>
<td>GPT-3</td>
<td>3.73</td>
<td>4.17</td>
<td>4.27</td>
</tr>
<tr>
<td>BART</td>
<td>4.13</td>
<td>3.81</td>
<td>4.15</td>
</tr>
<tr>
<td>BART, Predict</td>
<td>4.09</td>
<td>3.95</td>
<td>4.11</td>
</tr>
<tr>
<td>BART, Control</td>
<td>4.23</td>
<td>4.07</td>
<td>4.27</td>
</tr>
</tbody>
</table>

6. Conclusion + Future Directions
- **NLP:** semantics + new perspectives
- **Design + HCI:** implementation, user studies

- **Contradictions to the Premise:** (9%)
  - Insubstantial changes (26%)
  - Self-Contradictions (6%)
  - Hallucinations (2%)
- **Pragmatic Reconstruction**
  - Guided Seq. Generation
  - Persona Consistency