Have We Solved The **Hard** Problem? It’s Not **Easy**!

**Contextual Lexical Contrast as a Means to Probe Neural Coherence**

Wenqiang Lei, Yisong Miao, Runpeng Xie, Bonnie Webber, Meichun Liu, Tat-Seng Chua and Nancy Chen

**Introduction**

**Contextual Lexical Contrast (CLC)**

**Definition of CLC (a new NLP task):**
Two words are understood as contrast in order to understand the coherence of context.

**Example 1 (Pos.):** A *positive* attitude helps you relax and ace the exams, and a *negative* mental status will however make you nervous and sleepless.

**Example 2 (Neg.):** The reviewers are rather *positive* about this paper. They are nominating it for the Best Paper for its discovery of a *negative* finding that dispels conventional wisdom.

**Background**

- **Text Coherence**
- **Discourse (esp Comparison Relation)**
- **Humor and Contradiction Detection**

**Natural Language Processing (NLP)**

**Deep Learning**

**Cont2Lex Corpus**

- **Problem Formalization:** Given $w^+$ and $w^-$ in context $c$ (a sequence of words $w_1, w_2, \ldots, w_n$), a human (or a machine) needs to indicate a binary *tag* for CLC.

**Corpus Statistics:**
- Total 6,316 instances.
- Positive ratio: 35.7 (Adj. and Adv. have higher ratio)
- Inter-Annotator Agreement: 75.3%

**Table 1: Main Experiment (Acc Score)**

<table>
<thead>
<tr>
<th>Model</th>
<th>Glove</th>
<th>Word2Vec</th>
<th>FastText</th>
<th>ELMo</th>
<th>GPT</th>
<th>BERT</th>
<th>Majority</th>
</tr>
</thead>
<tbody>
<tr>
<td>BiLSTM</td>
<td>65.3</td>
<td>64.9</td>
<td>65.3</td>
<td>65.6</td>
<td>64.7</td>
<td>66.3</td>
<td></td>
</tr>
<tr>
<td>Attention</td>
<td>65</td>
<td>65.7</td>
<td>64.7</td>
<td>66.2</td>
<td>65.5</td>
<td>66.3</td>
<td></td>
</tr>
<tr>
<td>None</td>
<td>65.6</td>
<td>65.6</td>
<td>65.7</td>
<td>64.8</td>
<td>66.9</td>
<td>66.4</td>
<td>69.1</td>
</tr>
</tbody>
</table>

**Table 2: Out-of-context Lexical Contrast (Acc Score)**

**Possible Confounding Factors:**
1. $w^+$ and $w^-$ in the same sentence ($S$)
2. $w^+$ and $w^-$ is surrounded by repetitive words ($R$)

<table>
<thead>
<tr>
<th></th>
<th>S</th>
<th>$\neg$S</th>
<th>R</th>
<th>$\neg$R</th>
<th></th>
<th>S</th>
<th>$\neg$S</th>
<th>R</th>
<th>$\neg$R</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Glove</td>
<td>+4.2</td>
<td>-2.0</td>
<td>+7.2</td>
<td>-3.1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BERT.Lex</td>
<td>+3.6</td>
<td>-0.1</td>
<td>+5.0</td>
<td>-0.4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BERT</td>
<td>+10.3</td>
<td>+1.5</td>
<td>+14.9</td>
<td>+0.3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Majority</td>
<td>57.1</td>
<td>69.9</td>
<td>53.7</td>
<td>70.4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Table 3: Fine-grained Acc Scores**

**Conclusion**

- CLC is a challenging semantic representation task.
- Contextual embeddings perform better, but their advantage are largely due to capturing surface patterns.