ELMo Tasks: Sentiment

- this movie was great! would watch again
- the movie was gross and overwrought, but I liked it
- this movie was not really very enjoyable

Why are context-dependent embeddings useful here?

Peters et al. (2018)

ELMo Tasks: NER

- Barack Obama will travel to Hangzhou today for the G20 meeting.

Why are context-sensitive embeddings useful here?

Peters et al. (2018)

ELMo Tasks: SRL

- Housing starts are expected to quicken a bit from August's pace

Why are context-dependent embeddings useful here?

Peters et al. (2018)
**ELMo**

<table>
<thead>
<tr>
<th>TASK</th>
<th>PREVIOUS SOTA</th>
<th>OUR BASELINE</th>
<th>ELMo + BASELINE</th>
<th>INCREASE (ABSOLUTE/RELATIVE)</th>
</tr>
</thead>
<tbody>
<tr>
<td>SQuAD</td>
<td>Liu et al. (2017)</td>
<td>84.4</td>
<td>81.1</td>
<td>85.8</td>
</tr>
<tr>
<td>SNLI</td>
<td>Chen et al. (2017)</td>
<td>88.6</td>
<td>88.0</td>
<td>88.7 ± 0.17</td>
</tr>
<tr>
<td>SRL</td>
<td>He et al. (2017)</td>
<td>81.7</td>
<td>81.4</td>
<td>84.6</td>
</tr>
<tr>
<td>Coref</td>
<td>Lee et al. (2017)</td>
<td>67.2</td>
<td>67.2</td>
<td>70.4</td>
</tr>
<tr>
<td>NER</td>
<td>Peters et al. (2017)</td>
<td>91.93 ± 0.19</td>
<td>90.15</td>
<td>92.22 ± 0.10</td>
</tr>
<tr>
<td>SST-5</td>
<td>McCann et al. (2017)</td>
<td>53.7</td>
<td>51.4</td>
<td>54.7 ± 0.5</td>
</tr>
</tbody>
</table>

- Training: 1B words of text taken from MT data (news, etc.)
- SST5 = sentiment, SQuAD = QA (coming later)
- Large gains across many tasks

Peters et al. (2018)

**ELMo**

- Highly flexible: can use as an embedding layer for nearly any task
- Can be big and slow: big LSTMs can suck up GPU memory
- Strictly better than word2vec if you can afford it
- AllenNLP: calling ELMo is just one line, plenty of models for different tasks built on top of it

Peters et al. (2018)