Generalized Supervised Attention for Text Generation

Speaker: Yixian Liu

Advicer: Kewei Tu
Alignments in Text Generation

Target Sentence:

“From among them, pick out 50 for submission to an assessment committee to assess.”

➔ Ideal alignments (10) costly or inaccessible
➔ Quasi alignments (8), (10) accessible
➔ Relevant alignments (4), (10), (12), (14) hard to indicate
Generalized Supervised Attention (GSA)
Attention Supervision from Quasi Alignment

- **Attention supervision**
  \[
  \tilde{\alpha}_t = [0, 0, 1, 0, 0, 1]
  \]

- **Candidate set**
  A set containing all indexes of quasi alignments of the target word
  \[
  C(y_t) = \{2, 5\}
  \]

- **Objective function**
  \[
  \text{Loss} = \ell_{\text{mle}}(x, y) + \lambda \sum_t \Delta(\alpha_t, \hat{\alpha}_t)
  \]

---

**ROTOWIRE Dataset (Wiseman, 2017)**

<table>
<thead>
<tr>
<th>TEAM</th>
<th>Win</th>
<th>LOSS</th>
<th>PTS</th>
<th>FG_PCT</th>
<th>RB</th>
<th>AST</th>
</tr>
</thead>
<tbody>
<tr>
<td>Toronto Raptors</td>
<td>11</td>
<td>12</td>
<td>103</td>
<td>49</td>
<td>47</td>
<td>27</td>
</tr>
<tr>
<td>Cleveland Cavallers</td>
<td>7</td>
<td>15</td>
<td>95</td>
<td>43</td>
<td>33</td>
<td>20</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>PLAYER</th>
<th>AS</th>
<th>RB</th>
<th>PT</th>
<th>FG</th>
<th>FGA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Terrence Ross</td>
<td>5</td>
<td>2</td>
<td>27</td>
<td>8</td>
<td>11</td>
</tr>
<tr>
<td><strong>James Johnson</strong></td>
<td><strong>4</strong></td>
<td><strong>17</strong></td>
<td><strong>23</strong></td>
<td><strong>9</strong></td>
<td><strong>11</strong></td>
</tr>
<tr>
<td>LeBron James</td>
<td>2</td>
<td>9</td>
<td>21</td>
<td>8</td>
<td>12</td>
</tr>
<tr>
<td>Shawn Marion</td>
<td>2</td>
<td>3</td>
<td>19</td>
<td>7</td>
<td>15</td>
</tr>
<tr>
<td><strong>James Jones</strong></td>
<td><strong>5</strong></td>
<td><strong>5</strong></td>
<td><strong>10</strong></td>
<td><strong>5</strong></td>
<td><strong>11</strong></td>
</tr>
</tbody>
</table>

... was **James Johnson**, who added **11** ...
Cross Entropy (CE) and Summation Cross Entropy (SCE)

\[ \tilde{\alpha}_t = [0, 0, 1, 0, 0, 1] \]

- **CE**
  \[ \Delta(\alpha_i, \hat{\alpha}_i) = -\sum_{j=1}^{m} \hat{\alpha}_{i,j} \times \log \alpha_{i,j} \]

- **SCE**
  \[ \Delta(\alpha_t, \tilde{\alpha}_t) = \begin{cases} 
  0, \text{ if } \tilde{\alpha}_t = 0 \\
  -\log(\langle \alpha_t, \tilde{\alpha} \rangle), \text{ otherwise} 
\end{cases} \]

SCE loss reduces the effect of incorrect alignments in the candidate set, which promotes the true source word.
Supervised Multiple Attention (SMA)

- **MA**

\[ c_{t}^{(k)} = \text{attn}(s_{t}, H; \theta_{k}) \text{ for } k = 1, \ldots, K \]

\[ c_{t} = G(c^{(1)}, \ldots, c^{(K)}) \]

- **SMA**

The first channel of MA is supervised.

\[ \text{Loss} = \ell_{mle}(x, y) + \lambda \sum_{t} \Delta(\alpha_{t}^{(1)}, \hat{\alpha}_{t}) \]

SMA balances the supervised attention and unsupervised attention.
Structured Data to Text Generation

Task

Generate an abstract from a graph

Aligner

- node: string match
- edge: rule

Model

- encoder: Graph Transformer
- decoder: RNN

(Koncel-Kedziorski, 2019)

We evaluate MODEL1 on TASK1. MODEL1 outperforms MODEL2 by 15% on TASK1.
**AMR to Text Generation**

- **Task**
  Generate a sentence from an AMR graph

- **Aligner**
  word lamma match

- **Model**
  fine tuned GPT-2
  
  (Damonte, 2019)

---

**AMR dataset (Banarescu, 2013)**

```
arg1
  arg1-of
    arg0-of
    arg3
    arg2
    arg2
  arg1
  arg0
  arg3
```

From among them, pick out 50 for submission to an assessment committee to assess.
gunshots were fired at rapper lil wayne's tour bus early sunday in atlanta. no one was injured in the shooting, and no arrests have been made, atlanta police spokesman elizabeth espy said. police are still looking for suspects. officers were called to a parking lot in atlanta's buckhead neighborhood, espy said. they arrived at 3:25 a.m. and located two tour buses that had been shot multiple times. the drivers of the buses said the incident occurred on interstate 285 near interstate 75, espy said. witnesses provided a limited description of the two vehicles suspected to be involved: a "corvette style vehicle" and an suv. lil wayne was in atlanta for a performance at compound nightclub saturday night. cnn's carma hassan contributed to this report.
## Result

<table>
<thead>
<tr>
<th>Approach</th>
<th>Data-to-text</th>
<th>AMR-to-text</th>
<th>Summarization</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>BLEU</td>
<td>METEOR</td>
<td>ROUGE-L</td>
</tr>
<tr>
<td>UA</td>
<td>14.30</td>
<td>18.80</td>
<td>27.94</td>
</tr>
<tr>
<td>SA-CE</td>
<td>15.21</td>
<td>19.47</td>
<td>28.31</td>
</tr>
<tr>
<td>SA-SCE</td>
<td>15.49</td>
<td>19.80</td>
<td>28.62</td>
</tr>
<tr>
<td>SMA-SCE</td>
<td><strong>15.51</strong></td>
<td><strong>19.88</strong></td>
<td><strong>29.00</strong></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Approach</th>
<th>A.C.</th>
<th>M.C.</th>
<th>M.S.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data2text</td>
<td>13.21%</td>
<td>9.41%</td>
<td>2</td>
</tr>
<tr>
<td>AMR2text</td>
<td>46.64%</td>
<td>27.11%</td>
<td>2.93</td>
</tr>
<tr>
<td>Text Sum.</td>
<td>83.53%</td>
<td>78.10%</td>
<td>8.98</td>
</tr>
</tbody>
</table>

1. Supervised attention is better than unsupervised attention

2. SCE outperforms CE with higher M.S.

A.C. Alignment Coverage

M.C. Multi-alignment Coverage

M.S. Average Multi-alignment Size
**Analysis of Attention by SCE**

SCE loss reduces the effect of incorrect alignments in the candidate set, which promotes the true source word.
## Robustness analysis

<table>
<thead>
<tr>
<th>Error Rate</th>
<th>Data-to-text</th>
<th>AMR-to-text</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>BLEU</td>
<td>METEOR</td>
</tr>
<tr>
<td>0%</td>
<td>15.51</td>
<td>19.88</td>
</tr>
<tr>
<td>5%</td>
<td>14.93</td>
<td>19.68</td>
</tr>
<tr>
<td>10%</td>
<td>14.49</td>
<td>19.00</td>
</tr>
<tr>
<td>20%</td>
<td>14.36</td>
<td>18.99</td>
</tr>
<tr>
<td>33%</td>
<td>14.30</td>
<td>19.18</td>
</tr>
<tr>
<td>UA</td>
<td>14.30</td>
<td>18.80</td>
</tr>
</tbody>
</table>

GSA is robust to alignment errors.
Thank You