Deep Inside-outside Recursive Autoencoder with All-span Objective Function

Ruyue Hong Jiong Cai Kewei Tu
School of Information Science and Technology, ShanghaiTech University

Motivation
The original objective function of DIORA is constrained to utilize leaf-level information only. However, higher-level spans also embody meaningful information which could be utilized.

Proposed Solution
We proposed a new objective function that utilizes information among all-level spans and assigns weights of spans based on scores.

Original Objective Function
\[ L_s = \sum_{i \in \Pi(x)} \max \left(0, 1 - \bar{b}(i) \cdot \bar{a}(i) + \bar{a}(i) \cdot \bar{a}'(i)\right) \]

max-margin loss:
\[ L_s = \sum_{i \in \Pi(x)} \max \left(0, 1 - \bar{a}(i) \cdot \bar{a}(i) + \bar{a}(i) \cdot \bar{a}'(i)\right) \]

softmax loss:
\[ L_s = \sum_{i \in \Pi(x)} \log \left(\frac{\exp(\bar{a}(i) \cdot \bar{a}(i)) + \exp(\bar{a}(i) \cdot \bar{a}'(i))}{Z'(i) + \exp(\bar{a}(i) \cdot \bar{a}'(i))}\right) \]

\( \Phi(x) \) denotes the set of leaf-spans in the sentence \( x \), \( \mathcal{N}(i) \) denotes the set of negative samples for span \( i \).

Empirical Results
Experiment results show that our new training objective performs well on datasets of two languages: English and Japanese. And empirically show that our method achieves improvement in parsing accuracy over the original DIORA.

All-span Objective Function
\[ L_s = n \cdot \sum_{i \in \Pi(x)} \frac{w_f(i)}{\sum_{j \in \Pi(x)} w_f(j)} \cdot L_i \]

\( w_f(i) = \exp(\frac{\bar{a}(i)}{m_i}) \)

\( w_f(i) = \exp(\frac{\bar{a}(i)}{n - m_i + 1}) \cdot \bar{a}(i) \)

\( \Pi(x) \): the set of all spans in the sentence \( x \).

\( L_i \): the loss of span \( i \).

\( w_f(i) \): weight based on inside scores.

\( w_f(i) \): weight based on outside scores.

\( m_i \): the length of span \( i \).

\( n \): the length of the sentence.

Experiments and Analysis

Dataset: PTB corpus (Marcus et al., 1993) for English.

Experimental results on English PTB.

<table>
<thead>
<tr>
<th></th>
<th>( F_{1-10} )</th>
<th>( F_{1-100} )</th>
<th>( F_{1-all} )</th>
<th>( F_{1-all} )</th>
<th>( F_{1-all} )</th>
<th>( F_{1-all} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>DIORA</td>
<td>55.15 ± 0.86</td>
<td>56.61</td>
<td>49.63 ± 0.66</td>
<td>45.02</td>
<td>60.01 ± 0.40</td>
<td>60.64</td>
</tr>
<tr>
<td>DIORA-all</td>
<td>59.18 ± 1.71</td>
<td>63.64</td>
<td>57.03 ± 4.20</td>
<td>50.63</td>
<td>61.23 ± 2.31</td>
<td>63.33</td>
</tr>
<tr>
<td>Upper Bound</td>
<td>75.15</td>
<td>76.96</td>
<td>80.54</td>
<td>85.34</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 1: Experimental results on English PTB.

<table>
<thead>
<tr>
<th></th>
<th>( F_{1-10} )</th>
<th>( F_{1-100} )</th>
<th>( F_{1-all} )</th>
<th>( F_{1-all} )</th>
<th>( F_{1-all} )</th>
<th>( F_{1-all} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>DIORA</td>
<td>49.31 ± 0.45</td>
<td>49.73</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DIORA-all</td>
<td>49.18 ± 2.10</td>
<td>58.92</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Dataset: KTB corpus (Butler et al., 2012) for Japanese.

Experimental results on Japanese KTB.

<table>
<thead>
<tr>
<th></th>
<th>( F_{1-10} )</th>
<th>( F_{1-100} )</th>
<th>( F_{1-all} )</th>
<th>( F_{1-all} )</th>
<th>( F_{1-all} )</th>
<th>( F_{1-all} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>DIORA</td>
<td>39.33 ± 2.92</td>
<td>42.81</td>
<td>29.91 ± 4.29</td>
<td>32.33</td>
<td>44.02 ± 5.02</td>
<td>49.18</td>
</tr>
<tr>
<td>DIORA-all</td>
<td>43.30 ± 5.18</td>
<td>47.73</td>
<td>33.00 ± 2.71</td>
<td>36.93</td>
<td>47.09 ± 1.79</td>
<td>49.17</td>
</tr>
<tr>
<td>Upper Bound</td>
<td>61.41</td>
<td>62.53</td>
<td>67.25</td>
<td>67.32</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 2: Experimental results on Japanese KTB.

- With punctuation: DIORA-all consistently outperforms DIORA especially on long spans.
- Without punctuation: the two methods have similar accuracy except for an outlier at length range [31-35].