Overview

We develop an effective inference procedure for the Choe and Charniak (2016) LSTM language model over linearized parse trees, achieving state-of-the-art single-model performance on the Penn Treebank.

Generative Model

Disparities between the log-probabilities of structural vs. word generation actions cause standard beam search to fail.

Word-Level Search with Fast-Track Candidates

Grouping candidates by the current word addresses the imbalance in probabilities.

OPEN Action Pruning

Using the last 2 actions and the next word, we can prune 70% of OPEN actions with negligible effect on F1.

Final Results

<table>
<thead>
<tr>
<th>Parser</th>
<th>F1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vinyals et al. (2015)</td>
<td>88.3</td>
</tr>
<tr>
<td>Cross and Huang (2016)</td>
<td>91.3</td>
</tr>
<tr>
<td>Dyer et al. (2016)</td>
<td>91.7</td>
</tr>
<tr>
<td>Stern et al. (2017)</td>
<td>91.79</td>
</tr>
<tr>
<td><strong>Our Best Result</strong></td>
<td>92.56</td>
</tr>
<tr>
<td><strong>Our Best Result (with pruning)</strong></td>
<td>92.53</td>
</tr>
<tr>
<td>Vinyals et al. (2015) (ensemble)</td>
<td>90.5</td>
</tr>
<tr>
<td>Choe and Charniak (2016) (rerank)</td>
<td>92.6</td>
</tr>
<tr>
<td>Dyer et al. (2016) (rerank)</td>
<td>93.3</td>
</tr>
<tr>
<td>Fried et al. (2017) (ensemble, rerank)</td>
<td>94.25</td>
</tr>
</tbody>
</table>