Motivation

Small mistakes (typos and word repetitions) can be problematic for syntactic parsers. In such cases, a parser should produce an analysis for the intended sentence, rather than for the given literal sentence.

The linguistic depth of broad-coverage grammars, like HPSG (Pollard and Sag, 1987), could be exploited for error correction. Since common grammar parsers do not allow cross-sentence comparison of parse tree probabilities, we evaluate the treelet model (Pauls and Klein, 2012) for estimating generative probabilities for HPSG parsing with error correction.

Treelet model

The treelet model is proposed in Pauls and Klein (2012):

\[
p(T) = \Pi_{\eta \in \mathcal{TR}_P}(c^\eta \mid \eta)
\]

where \( T \) is a constituency tree consisting of context-free rules of the form \( r = P \rightarrow C_1 \cdots C_p \), \( P \) is the parent symbol of the rule \( r \), \( C_1, \ldots, C_p \) are its children and \( \eta \) is the context.

The context for non-terminal productions includes parent node \( P \), grandparent node \( P' \) and the rule \( r' \) that generated \( P \); while for terminal (lexical) productions the context covers \( P \), the sibling \( R \) immediately to the right of \( P \), parent rule \( r \) and the previous two words \( w_{-1} \) and \( w_{-2} \) (see figures).

Treelet model for parse selection

Goal: compare the treelet model with the PCFG model on the parse selection problem

We compute how many times either model chooses the gold-standard analysis among the up to 500 analyses generated by the PET parser.

Number and percentage of sentences for which treelet, PCFG and random choice models scored the gold-standard parse tree higher than other analyses produced by the PET parser:

Upper bound: 12,311 sent.; 100%

PET: 4,487 sent.; 36.45%

PCFG: 2,305 sent.; 23.60%

Random: 621 sent.; 5.04%

Error analysis

The sentences on which the treelet model outperforms the PCFG model are:

- longer;
- contain slightly more coordination structures;
- most of the vocabulary has been seen by the model during training.

Future work

- enhancement of the treelet model implementation with the transformations suggested in Pauls and Klein (2012);
- supplying the system with a large dictionary of named entities;
- building a system for HPSG-parsing with error correction.

Acknowledgements

We are grateful to Lilja Óvrelid and Stephan Oepen for helpful feedback and discussions, as well as anonymous reviewers for their comments. The first author is funded by the Norwegian Research Council through its WeSearch project.

Conclusions

- the treelet model outperforms the random and PCFG models for the parse selection problem;
- the treelet model appears to perform better than the random, trigram and PCFG models on the NUS corpus data;
- differences in the performance of the treelet, PCFG and trigram models are not statistically significant on the Wikipedia data.

Treelet model for scoring parse trees of erroneous and corrected sentences

Goal: evaluate the treelet’s ability to assign a higher probability to the corrected sentence than to the erroneous one in comparison to the PCFG and SRILM trigram models.

The table demonstrates the number of sentences for which the treelet, PCFG and random choice models:

<table>
<thead>
<tr>
<th>Model</th>
<th>Corrected</th>
<th>Equal</th>
<th>Original</th>
<th>Corrected</th>
<th>Equal</th>
<th>Original</th>
</tr>
</thead>
<tbody>
<tr>
<td>NUS corpus</td>
<td>2,223</td>
<td>0</td>
<td>4,604</td>
<td>0</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td>Train</td>
<td>1,449</td>
<td>11</td>
<td>1,884</td>
<td>994</td>
<td>1,726</td>
<td></td>
</tr>
<tr>
<td>PCFG</td>
<td>1,304</td>
<td>11</td>
<td>1,335</td>
<td>995</td>
<td>1,773</td>
<td></td>
</tr>
<tr>
<td>Trigram</td>
<td>1,240</td>
<td>80</td>
<td>1,732</td>
<td>1,294</td>
<td>1,578</td>
<td></td>
</tr>
<tr>
<td>Random</td>
<td>1,112</td>
<td>1,111</td>
<td>2,302</td>
<td>2,302</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>