Parameter sharing between dependency parsers for related languages

Miryam de Lhoneux, Johannes Bjerva, Isabelle Augenstein and Anders Søgaard

Parsing Architecture

Configuration:

STACK

BUFFER

the brown fox

jumped root

Transitions:

LEFT–ARC

RIGHT–ARC

SHIFT

SWAP

Scoring:

(score(LEFT−ARC),score(RIGHT−ARC),score(SHIFT), score(SWAP))

State

Word

MLP

LSTM f

LSTM b

Character

Sharing strategies

soft sharing (hard) sharing not sharing

Not sharing

MLP or word/character lookup and BiLSTM parameters not shared

Hard sharing

MLP or word/character lookup and BiLSTM parameters shared

Soft sharing

Sharing + concatenating a language embedding to the configuration vector, or word or character vectors at the input of the BiLSTM.

We test the \(3^3 = 27\) combinations on 5 language pairs.

Results

<table>
<thead>
<tr>
<th>ISO</th>
<th>Lang</th>
<th>Tokens</th>
<th>Family</th>
<th>Word order</th>
</tr>
</thead>
<tbody>
<tr>
<td>en</td>
<td>English</td>
<td>208,812</td>
<td>Semitic</td>
<td>VSO</td>
</tr>
<tr>
<td>he</td>
<td>Hebrew</td>
<td>163,005</td>
<td>Semitic</td>
<td>SVO</td>
</tr>
<tr>
<td>et</td>
<td>Estonian</td>
<td>62,254</td>
<td>Finnic</td>
<td>SVO</td>
</tr>
<tr>
<td>fi</td>
<td>Finnish</td>
<td>67,254</td>
<td>Finnic</td>
<td>SVO</td>
</tr>
<tr>
<td>hr</td>
<td>Croatian</td>
<td>109,965</td>
<td>Slavic</td>
<td>SVO</td>
</tr>
<tr>
<td>ru</td>
<td>Russian</td>
<td>90,170</td>
<td>Slavic</td>
<td>SVO</td>
</tr>
<tr>
<td>it</td>
<td>Italian</td>
<td>113,825</td>
<td>Romance</td>
<td>SVO</td>
</tr>
<tr>
<td>es</td>
<td>Spanish</td>
<td>154,844</td>
<td>Romance</td>
<td>SVO</td>
</tr>
<tr>
<td>nl</td>
<td>Dutch</td>
<td>75,796</td>
<td>Germanic</td>
<td>No dom. order</td>
</tr>
<tr>
<td>no</td>
<td>Norwegian</td>
<td>76,622</td>
<td>Germanic</td>
<td>SVO</td>
</tr>
</tbody>
</table>

Las on the test sets of the best of 9 sharing strategies and the monolingual baseline. \(\delta\) is the difference between Ours and Mono.

Conclusions

• Generally, multi-task learning helps.
• Sharing the MLP parameters always helps. It helps to share MLP parameters when training a parser on a pair of related languages, and it also helps if the languages are unrelated.
• Sharing word and character parameters is differently helpful depending on the language.
• Sharing too many parameters does not help, when the languages are unrelated.

References