Character-Level Models (CLMs) vs Morphology in Semantic Role Labeling

Gözde Gül Şahin*, Mark Steedman

goze.de.guel@gmail.com, steedman@inf.ed.ac.uk

Ubiquitous Knowledge Processing (UKP) Lab, University of Edinburgh

*Work done partially in Istanbul Technical University (ITU) and University of Edinburgh

**Abstract**

The best model was the morphology-level model in all languages, BUT...

**The Role of Morphology in SRL**

Morphology is essential for semantic role labeling (SRL) but expensive, so we ask the following questions:

- Can character level models (CLMs) replace oracle (gold morphological analysis)?
- What are CLMs' limitations and strengths compared to oracle?

**Method**

- Words are decomposed into subwords
- Subwords are composed into word vectors
- Words are treated as a sequence of subwords
- Generated input vectors are fed to the sequence-labeling network
- For the sake of simplicity, the label with the highest probability is assigned to the input.

**Experiments**

- Dataset: CoNLL-09 dependency-based SRL shared task dataset for Czech, Spanish, Catalan, German and English and Free PropBank: Turkish PropBank [3] and Finnish PropBank

<table>
<thead>
<tr>
<th>word</th>
<th>char</th>
<th>char-trigram</th>
<th>gold morph</th>
<th>tags</th>
</tr>
</thead>
<tbody>
<tr>
<td>FINNISH</td>
<td>48.91</td>
<td>67.24</td>
<td>37.46</td>
<td>67.78</td>
</tr>
<tr>
<td>TURKISH</td>
<td>44.82</td>
<td>55.89</td>
<td>24.68</td>
<td>56.60</td>
</tr>
<tr>
<td>SPANISH</td>
<td>64.30</td>
<td>67.90</td>
<td>5.61</td>
<td>68.43</td>
</tr>
<tr>
<td>CATALAN</td>
<td>65.45</td>
<td>70.56</td>
<td>7.82</td>
<td>71.34</td>
</tr>
<tr>
<td>CZECH</td>
<td>63.58</td>
<td>74.04</td>
<td>16.45</td>
<td>74.98</td>
</tr>
<tr>
<td>GERMAN</td>
<td>54.78</td>
<td>63.71</td>
<td>16.29</td>
<td>65.56</td>
</tr>
</tbody>
</table>

**Table 1: Argument labeling F1 scores for each subword unit and language.**

- **Why does improvement over Word (IOW) range between 0%-38%??**
- **Why does improvement over Character (IOC) range between 2%-10%?**

**So the answers are...**

- For in-domain data, CLMs can not yet match the performance of MLMs, but surpass WLMs by a large margin
- Its shortcomings depend on the language type. The hard cases are: Derivational morphology and contextual ambiguity for agglutinative languages; tokens with many morphological tags in fusional languages.
- They perform better on out-of-domain data; when there is only access to predicted tags; and when a large enough training set is available. Targeted scores for long range dependencies are similar.
- They don't benefit as much from increasing of the model size and perform worse in case of small training data size.

**Acknowledgments**

Gözde Gül Şahin was funded by Tübitak (The Scientific and Technological Research Council of Turkey) 2214-A scholarship during her visit to University of Edinburgh. She was granted access to CoNLL-09 Semantic Role Labeling Shared Task data by Linguistic Data Consortium (LDC) in Fall 2015. This work was supported by ERC 66290 Advanced Fellowship GA 742137 SEMANTAX and a Google Faculty award to Mark Steedman.