Research Problem

• The SIGMORPHON 2019 Task 2 requires us to produce the lemma and morpho-syntactic description of each token in a sequence, for 107 treebanks.

• However, most treebanks are under-resourced, making it challenging to train deep neural models for them.

• We approach this task with a hierarchical neural conditional random field (CRF) model which predicts each coarse-grained feature (e.g., POS, Case, etc.) independently.

• To tackle the challenge of data-scarcity we propose a multi-lingual transfer training regime where we transfer from multiple related languages that share similar typology and/or orthography.

Model

Hierarchical neural model for contextual morphological analysis with independent CRF decoders for each coarse-grained feature \( F \) such as Number, Person etc. POS embeddings are concatenated to the word and char-level representation.

This model has \( |F|-1 \) decoders since POS tagger is run separately as a prior step. MDCRF refers to the above model without POS embeddings having all \( |F| \) decoders.

Results

• Our system achieves an average improvement of +15.30 (accuracy) and +4.93 (F1) over the provided baseline. (McCarthy et al., 2019).

• We compare our model for bi-lingual transfer with previous baselines:

<table>
<thead>
<tr>
<th>Language</th>
<th>Model</th>
<th>tp-class-1/00</th>
<th>tp-class-1/00</th>
</tr>
</thead>
<tbody>
<tr>
<td>BUBA</td>
<td>MDCRF + POS + SUFFIXES</td>
<td>68.60</td>
<td>65.70</td>
</tr>
<tr>
<td></td>
<td>(McCarthy et al., 2018)</td>
<td>68.48</td>
<td>65.68</td>
</tr>
<tr>
<td></td>
<td>(Chaudhary et al., 2017)</td>
<td>68.48</td>
<td>65.68</td>
</tr>
<tr>
<td>FSWI</td>
<td>MDCRF + POS + SUFFIXES</td>
<td>75.32</td>
<td>76.50</td>
</tr>
<tr>
<td></td>
<td>(McCarthy et al., 2018)</td>
<td>75.32</td>
<td>76.50</td>
</tr>
<tr>
<td></td>
<td>(Chaudhary et al., 2017)</td>
<td>75.32</td>
<td>76.50</td>
</tr>
</tbody>
</table>

Q. Why does adding POS help?

• Removing POS gives a drop in accuracy which is significant for low-medium resource languages: Marathi (-6.12 acc), Ukrainian (-3.57 acc).

• Number of errors per feature \( F \): POS helped reduced Gender errors in Marathi.

Eg. For some words gender may be inferred from inflectional form, but for others such as “(price.N.FEM.SG.ACC)” the traditional female suffix is not present, in which case having the POS information helps.

Q. What is the model learning?

• Character-level attention maps for three typologically different languages:

- **INDONESIAN**
- **MARATHI**
- **BELARUSIAN**

Marathi and Belarusian display morphological inflections as suffix.

Indonesian displays inflections in the form of prefix, suffix and circumfix.

Q. Does time-depth matter for transfer learning?

• **Time-depth** is period of time that has elapsed since all languages in the group were a single language.

• We study the following clusters: Indo-Aryan, Slavic and Semitic.

- Transfer helps most for Slavic (+2.9 accuracy), and next for Indo-Aryan cluster (+0.32 accuracy). There is a negative effect for the Semitic cluster (-0.0176 accuracy).

- Indo-Aryan and Slavic have shallower time-depths: < 1000 years ⇒ better transfer!