Overcoming Early Saturation on Low-Resource Languages in Multilingual Dependency Parsing

Jiannan Mao†, Chenchen Ding‡, Hour Kaing†,
Hideki Tanaka‡, Masao Utiyama‡, Tadahiro Matsumoto†

†Gifu University, Gifu, Japan
‡National Institute of Information and Communications Technology, Kyoto, Japan
Outline

• Overview
• Background
• Investigation
• Experiments
• Conclusion and Future Work
• Dependency parsing:
  1. linguistic analysis technique
  2. uncover grammatical relationships (words in a sentence).

• Large treebanks: efficiently.

• For low-resource languages: treebanks are small/unavailable.

Multilingual Parsing

I have a dog.
Multilingual Parsing and UDify

- A single model: parsing different languages.
- The lack of data: cross-lingual information.
  - A multilingual multi-task parser.
  - Exhibits strong and consistent performance in all UD treebanks.
  - Early saturation occurs in some low-resource languages.
Contributions:

- on multiple low-resource languages using data augmentation methods.
- the unlabeled attachment score (UAS), enhance stability:
- Robustness of multilingualism processing is still retained.
UDify

- Lemmas, POS tags, and dependency structures.
- Finetuned on multilingual BERT.
- No language tag

English raw sentence input example: The best optimizer is grad student descent
An unsupervised algorithm for dependency learning (Unsupervised-Dep).

Constructs the tree
1. a dynamic programming method (CYK chart)
2. the complete-link and complete-sequence

Considering the time complexity of N-gram, focus on the bi-gram.
Unsupervised Dependency Learning: **bi-gram**

- Dependency relations define pair directions:
  \[(w_i \rightarrow w_j), (w_i \leftarrow w_j).\]

- Calculated using the Inside-Outside algorithm.

- Tree construction determined by Viterbi algorithm to ensure maximum probability.
UDify with Data Augmentation - Training

1. Feed the $D_{\text{train}}$, into trained-UDify

2. Statistical computations are performed on $\text{DEP}_{\text{arc}}$

3. Unsupervised-Dep:
   $P(w_i \rightarrow w_j), P(w_i \leftarrow w_j)$ and $D_{\text{train}}$

4. Obtain the re-estimated probabilities
   $P(w_i \rightarrow w_j)'$ and $P(w_i \leftarrow w_j)'$
Find the optimal structure of $D_{test}$ by $P(w_i \rightarrow w_j)'$, $P(w_i \leftarrow w_j)'$ and Viterbi.

Retain information other than $\text{DEP}_{arc}$ of the parser result from UDify.

Construct the $D_{test}$ in the UD treebank format.

Train a new UDify.
On Few- and Zero-Shot Languages

- Early saturation in the accuracy of dependency parsing was observed.

- Unsupervised-Dep data augmentation across multiple low-resource languages remains underexplored.
### Experiments

**Dataset and Setup**

<table>
<thead>
<tr>
<th>language(code)</th>
<th>#sent.(len.)</th>
<th>#train</th>
<th>#test</th>
</tr>
</thead>
<tbody>
<tr>
<td>Armenian(hy)</td>
<td>2.4(8.2)</td>
<td>560</td>
<td>470</td>
</tr>
<tr>
<td>Belarusian(be)</td>
<td>2.0(9.0)</td>
<td>260</td>
<td>68</td>
</tr>
<tr>
<td>Hungarian(hu)</td>
<td>134.1(5.3)</td>
<td>910</td>
<td>449</td>
</tr>
<tr>
<td>Kazakh(kk)</td>
<td>1.7(8.2)</td>
<td>31</td>
<td>1,047</td>
</tr>
<tr>
<td>Lithuanian(lt)</td>
<td>236.7(5.6)</td>
<td>153</td>
<td>55</td>
</tr>
<tr>
<td>Marathi(mr)</td>
<td>1.5(10.0)</td>
<td>373</td>
<td>47</td>
</tr>
<tr>
<td>Tamil(ta)</td>
<td>13.7(7.7)</td>
<td>400</td>
<td>120</td>
</tr>
<tr>
<td>Breton(br)</td>
<td>18.2(9.5)</td>
<td>0</td>
<td>888</td>
</tr>
<tr>
<td>Faroese(fo)</td>
<td>1.3(8.1)</td>
<td>0</td>
<td>1,208</td>
</tr>
<tr>
<td>Tagalog(tl)</td>
<td>150.0(16.2)</td>
<td>0</td>
<td>55</td>
</tr>
<tr>
<td>Yoruba(yo)</td>
<td>9.7(8.1)</td>
<td>0</td>
<td>100</td>
</tr>
</tbody>
</table>

- **UDify(our)**: Reproduced UDify model.
- **Unsup**: UD 2.3+OPUS-mult_{300}, generated by Unsupervised-Dep.
- **Self**: UD 2.3+ OPUS-mult_{300}, the parsing results from Baseline.

OPUS-mutt: Raw data collected from various corpora.
## Experiments

### Dependency Task on the Low-Resource Languages

<table>
<thead>
<tr>
<th></th>
<th>hy</th>
<th>be</th>
<th>hu</th>
<th>kk</th>
<th>lt</th>
<th>mr</th>
<th>ta</th>
<th>br</th>
<th>fo</th>
<th>tl</th>
<th>yo</th>
<th>Few</th>
<th>Zero</th>
</tr>
</thead>
<tbody>
<tr>
<td>UDify(org)</td>
<td>85.6</td>
<td>91.8</td>
<td>89.7</td>
<td>74.8</td>
<td>79.1</td>
<td><strong>79.4</strong></td>
<td>79.3</td>
<td>63.5</td>
<td>67.2</td>
<td>64.0</td>
<td>37.6</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>UDify(our)</td>
<td>86.1</td>
<td>92.1</td>
<td>89.8</td>
<td>76.0</td>
<td>79.4</td>
<td>74.3</td>
<td>80.8</td>
<td>69.2</td>
<td>72.0</td>
<td>78.4</td>
<td>39.4</td>
<td>84.0</td>
<td>67.1</td>
</tr>
<tr>
<td>Self</td>
<td>85.9</td>
<td><strong>92.5</strong></td>
<td>89.6</td>
<td><strong>76.2</strong></td>
<td>79.2</td>
<td>74.8</td>
<td><strong>81.2</strong></td>
<td>69.8</td>
<td><strong>72.5</strong></td>
<td>85.3</td>
<td>38.8</td>
<td>84.0</td>
<td>67.6</td>
</tr>
<tr>
<td>Unsup</td>
<td><strong>86.3</strong></td>
<td>92.4</td>
<td><strong>90.0</strong></td>
<td><strong>76.2</strong></td>
<td><strong>79.5</strong></td>
<td>74.0</td>
<td>80.5</td>
<td><strong>72.7</strong></td>
<td>71.9</td>
<td><strong>88.0</strong></td>
<td><strong>39.6</strong></td>
<td><strong>84.2</strong></td>
<td><strong>68.7</strong></td>
</tr>
</tbody>
</table>

UAS for few- and zero-shot languages obtained using different methods.

**few-shot languages**: contain a little of training data in UD treebanks.

**zero-shot languages**: do not contain any training data in UD treebanks.
Other Tasks and Languages

<table>
<thead>
<tr>
<th></th>
<th>Zero-shot</th>
<th>Other</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>UAS</td>
<td>Rest</td>
</tr>
<tr>
<td>UDify (our)</td>
<td>67.1</td>
<td>55.6</td>
</tr>
<tr>
<td>Self</td>
<td>67.6</td>
<td>56.3</td>
</tr>
<tr>
<td>Unsup</td>
<td><strong>68.7</strong></td>
<td><strong>59.0</strong></td>
</tr>
</tbody>
</table>

UD scores on selected zero-shot and other languages obtained by different methods.

Difference in UAS on all test treebanks. X-axis: sorted order of treebank training sets from smallest to largest.
• Employed data augmentation through unsupervised learning.

• Overcome early saturation in parsing accuracy among low-resource languages.

• Future Work:
  • exploring additional factors
  • using the latest UD treebanks
Thank you for Listening!