Phylogeny-Inspired Adaptation of Multilingual Models to New Languages

Fahim Faisal
Antonios Anastasopoulos
ffaisal@gmu.edu
https://nlp.cs.gmu.edu/
Going beyond the top-100 languages

Make MLMs highly multilingual

Train them on 100 languages

Apply them on the other 6400 languages

Apply them on the *other* 6400 languages
Getting a LM for a new language

**Option A:** Train (a monolingual) one from scratch

**Option B:** Just use mBERT (zero-shot)

**Option C:** Continue training mBERT with same unsupervised objective

**Option D:** Adapters (Pfeiffer et al. 2020)
Revisiting Adapters
Revisiting Adapters
Revisiting Adapters
Revisiting Adapters

Easy zero-shot adaptation to new languages at a low cost (additional parameters)

Avoids catastrophic forgetting

Performance comparable to full-model fine-tuning

Can we do better?
Follow Phylogeny for Parameter Sharing
Follow Phylogeny for Parameter Sharing

For Dutch input
Follow Phylogeny for Parameter Sharing

For Bengali input
Experimental setup

G: Tupari
- L: Makuráp
- L: Akuntsu
- L: Munduruku
- L: Tupinambá
- L: Kaapor
- L: Mbya Guaraní
- L: Guajájara
- L: Simba Guaraní
- L: Guaraní

G: Munduruku

G: Tupi Guaraní

G: Tarahumaran
- L: Rarámuri
- L: Yaqui
- L: Mayo Tepehuan
- L: Southern Tepehuan
- L: O’odham
- L: Northern Tepehuan
- L: Huichol
- L: Cora
- L: Nahuatl

G: Tepiman

G: Corachol

G: Aztecan

F: Uto-Aztecan

F: Tupian
Results

DEPENDECE PARSING

<table>
<thead>
<tr>
<th>Language</th>
<th>[T]</th>
<th>[LT]</th>
<th>[FGLT]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Germanic</td>
<td>70.6</td>
<td>69.2</td>
<td>72.3</td>
</tr>
<tr>
<td>Uralic</td>
<td>48.3</td>
<td>51.4</td>
<td>58.3</td>
</tr>
<tr>
<td>Tupian</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Language Counts:
- Germanic: 12
- Uralic: 11
- Tupian: 8
Results on unseen languages

**DEPENDECY PARSING**

Much larger improvements for *new, unseen* languages
Results on unseen languages

Much larger improvements for *new, unseen* languages

You’re just using more parameters!
Ablations: Parameters

DEPENDENCY PARSING ON URALIC LANGS

Even constraining to the same number of parameters, still improvements!
Is it language sharing or network depth?
Ablations: Upsampling

<table>
<thead>
<tr>
<th>Language</th>
<th>SME</th>
<th>KOI</th>
<th>MYV</th>
<th>OLO</th>
<th>MDF</th>
<th>SMS</th>
<th>KPV</th>
<th>KRL</th>
</tr>
</thead>
<tbody>
<tr>
<td>FGLT</td>
<td>31.6</td>
<td>38.2</td>
<td>79.6</td>
<td>42.8</td>
<td>44.5</td>
<td>22.8</td>
<td>35.6</td>
<td>65.8</td>
</tr>
<tr>
<td>FGLT-[U]</td>
<td>45.2</td>
<td>44.1</td>
<td>53.8</td>
<td>55.9</td>
<td>53.8</td>
<td>23.0</td>
<td>42.4</td>
<td>69.6</td>
</tr>
</tbody>
</table>

Upsampling by simple repeating sentences does better.
• Adapter-based approach to leverage language phylogenetic information for better cross-lingual adaptation.

• Exact same parameter count but smaller adapters with parameter sharing across related language improves performance in true-zero-resource scenarios.
Code & Dataset

https://github.com/ffaisal93/adapt_lang_phylogeny

FAHIM FAISAL (ffaisal@gmu.edu)

Antonios Anastasopoulos (antonis@gmu.edu)