Cross-lingual Dependency Parsing with Unlabeled Auxiliary Languages

Wasi Ahmad, Zhisong Zhang, Xuezhe Ma, Kai-Wei Chang, and Nanyun Peng.
CoNLL, 2019
Motivation

Different languages have different properties (e.g., word order)

Improve transfer learning across languages (Learning language-agnostic representation)
Dependency Parsing

An encoder to produce contextualized representations

A decoder that makes (structured) predictions

Multilingual embeddings for the input sentence

I prefer the morning flight through Denver

Dependency Parser
Cross-lingual Dependency Parsing

Train

English Treebank

Parse

French Corpus

Aujourd'hui j'ai rencontré un accident.
J'ai besoin de prendre le vol.
Je ne pouvais pas déjeuner aujourd'hui à cause d'une réunion.
Main Idea

Train

English Treebank

Parse

French Corpus

Russian Corpus

Aujourd'hui j'ai rencontré un accident
J'ai besoin de prendre le vol
Je ne pouvais pas déjeuner aujourd'hui à cause d'une réunion

У меня сильная головная боль
Они прекрасно проводят время вместе
Они скоро поженятся
Main Idea

• Use unlabeled corpora of auxiliary languages

• Adversarial training to learn language-agnostic representation
  • **Discriminator**: predicts language label
Training Procedure

Step 1. Warm-start the parser

- Mini-batch training using source language treebank
- For k iterations

I prefer the morning flight through Denver

United canceled the morning flights to Houston

JetBlue canceled our flight this morning which was already late
Training Procedure

Step 2. Jointly train using auxiliary languages
- Train the parser on source language
- Adversarial training on both source and auxiliary languages

JetBlue canceled our flight this morning which was already late

Je ne pouvais pas déjeuner aujourd'hui à cause d'une réunion
Experiment Setup

Embedding

• Token embeddings
  • Multilingual Embeddings (MUSE) [Smith et al., 2017, Bojanowski et al., 2017]
  • Multilingual BERT (M-BERT) [Devlin et al., 2017]

• Part-of-speech embeddings

Parsers [Ahmad et al., 2019]

• Graph-based: Self-attentive-Graph
  • Multi-Head Self-Attention (order-free)

• Transition-based: RNN-StackPtr
  • BiLSTMs (order-dependent)
Experiment Setup

Single Source Transfer Parsing
- Train parser on one language
- Source language: English

Adversarial training
- Using a language pair (one source and one auxiliary language)
- Auxiliary languages are selected based on -
  - Covering different language families
  - Average distance between auxiliary language and all target languages
Experiment Setup

Datasets

• Universal Dependency Treebanks (v2.2)
• 1 Source language, 28 target languages
• 10 families

Evaluation

• Evaluate directly on the target languages (zero-shot)
• Metrics: UAS, LAS

<table>
<thead>
<tr>
<th>Language Families</th>
<th>Languages</th>
</tr>
</thead>
<tbody>
<tr>
<td>Afro-Asiatic</td>
<td>Arabic (ar), Hebrew (he)</td>
</tr>
<tr>
<td>Austronesian</td>
<td>Indonesian (id)</td>
</tr>
<tr>
<td>IE.Baltic</td>
<td>Latvian (lv)</td>
</tr>
<tr>
<td>IE.Germanic</td>
<td>Danish (da), Dutch (nl), English (en), German (de), Norwegian (no), Swedish (sv)</td>
</tr>
<tr>
<td>IE.Indic</td>
<td>Hindi (hi)</td>
</tr>
<tr>
<td>IE.Latin</td>
<td>Latin (la)</td>
</tr>
<tr>
<td>IE.Romance</td>
<td>Catalan (ca), French (fr), Italian (it), Portuguese (pt), Romanian (ro), Spanish (es)</td>
</tr>
<tr>
<td>IE.Slavic</td>
<td>Bulgarian (bg), Croatian (hr), Czech (cs), Polish (pl), Russian (ru), Slovak (sk), Slovenian (sl), Ukrainian (uk)</td>
</tr>
<tr>
<td>Korean</td>
<td>Korean (ko)</td>
</tr>
<tr>
<td>Uralic</td>
<td>Estonian (et), Finnish (fi)</td>
</tr>
</tbody>
</table>

Table 1: The selected 29 languages for experiments from UD v2.2 (Nivre et al., 2018).
Impact of Adversarial Training

Self-attentive-Graph Parser

Multilingual Word Embeddings

- Baseline
- French
- Russian

Multilingual BERT

- Baseline
- French
- Russian

x-axis = language labels
y-axis = performance_diff (model_trained_with_aux_lang, model_trained_on_src_lang)
Impact on Language Families

- Baseline
- Russian
- French
- German

IE.Slavic
IE.Romance
IE.Germanic
### Adversarial training (AT) vs. Multi-task learning (MTL)

<table>
<thead>
<tr>
<th>Adversarial Training (AT)</th>
<th>Multi-task Learning (MTL)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Encoder is trained to <strong>maximize</strong> the classifier’s loss</td>
<td>Encoder is trained to <strong>minimize</strong> the classifier’s loss</td>
</tr>
<tr>
<td>Learns to <strong>avoid</strong> distinguishable language features</td>
<td>Learns to <strong>retain</strong> distinguishable language features</td>
</tr>
<tr>
<td>Observes same amount of data</td>
<td>Observes same amount of data</td>
</tr>
</tbody>
</table>
Adversarial training (AT) vs. Multi-task learning (MTL)

x-axis = auxiliary language labels
y-axis = performance_diff (model_trained_with_aux_lang, model_trained_on_src_lang)
Language Test

- Tests how much encoders retain language information
- A multi-layer perceptron (MLP) as a 7-way classifier
  - Trained on the source and six auxiliary languages
- Metric: accuracy
Adversarial training (AT) vs. Multi-task learning (MTL) – Language Test

x-axis = auxiliary language labels
y-axis = language test performance
Conclusion

• Utilization of Unlabeled Language Corpora
  • Improves cross-lingual dependency parsing
• Adversarial training
  • To learn language-agnostic representation
• Comprehensive empirical study
• Future work
  • Multi-source Transfer Parsing

Source Code is Publicly Available
https://github.com/wasiahmad/cross_lingualParsing
References


Thank You
Impact of Auxiliary Languages

• Utilization of unlabeled auxiliary languages helps!

<table>
<thead>
<tr>
<th>Language</th>
<th>Multilingual Word Embeddings</th>
<th>Multilingual BERT</th>
</tr>
</thead>
<tbody>
<tr>
<td>En</td>
<td>65.92</td>
<td>73.34</td>
</tr>
<tr>
<td>En + Pt</td>
<td>66.40</td>
<td>73.47</td>
</tr>
<tr>
<td>En + Ru</td>
<td>66.53</td>
<td>73.88</td>
</tr>
<tr>
<td>En + De</td>
<td>66.41</td>
<td>73.92</td>
</tr>
<tr>
<td>En + Es</td>
<td>66.38</td>
<td>71.71</td>
</tr>
<tr>
<td>En + Fr</td>
<td>66.40</td>
<td>73.55</td>
</tr>
<tr>
<td>En + La</td>
<td>66.45</td>
<td>73.69</td>
</tr>
</tbody>
</table>