Combining Global Models for Parsing Universal Dependencies

Team C2L2 —
Tianze Shi, Felix G. Wu, Xilun Chen, Yao Cheng

Cornell University
Overview — Scope of Our System

What we did

- Projective Parsing
- Dependency Arc Labeling
- Delexicalized Parsing

What we didn’t do

- Word Segmentation
- Sentence Boundary Detection
- POS Tagging
- Morphology Analysis
- Non-projective Parsing
- Unlabeled data
Overview — Highlights

- argmax
  \[ y \in \mathcal{Y} \]
  - Global transition-based models

- Bi-LSTM-powered compact features

- Delexicalized syntactic transfer

- High efficiency, low resource demand

- Overall
  - 1st
    - Small Treebanks
    - Surprise Languages

- 2nd
Overview — System Pipeline

I. UDPipe Pre-process

II. Feature Extraction

III. Unlabeled Parsing

IV. Arc Labeling
I. UDPipe Pre-process

II. Feature Extraction

III. Unlabeled Parsing

IV. Arc Labeling

Raw Text

UDPipe

Sentence delimited & tokenized
<table>
<thead>
<tr>
<th>Languages</th>
<th>OOV rates ↓ (word)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ko – Korean</td>
<td>43.68%</td>
</tr>
<tr>
<td>la – Latin</td>
<td>41.22%</td>
</tr>
<tr>
<td>sk – Slovak</td>
<td>36.51%</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>Average</td>
<td>14.4%</td>
</tr>
</tbody>
</table>

* Measured on development set
I. UDPipe Pre-process

II. Feature Extraction

III. Unlabeled Parsing

IV. Arc Labeling

Bi-directional LSTM
I. UDPipe Pre-process

II. Feature Extraction

III. Unlabeled Parsing

IV. Arc Labeling

Bi-directional LSTM

Universal dependency parsing

Universal dependency parsing
I. UDPipe Pre-process

II. Feature Extraction

III. Unlabeled Parsing

IV. Arc Labeling

- Reparsing by Eisner’s (Sagae and Lavie, 2006)
  - Eisner’s
  - Arc-eager Global
  - Arc-hybrid Global

- Bi-LSTM features

NEW!!!
Global Transition-based Parsing

- $O(n^3)$ Exact decoders
- Arc-eager and Arc-hybrid systems
- Large-margin global training
- Dynamic programming (Huang and Sagae, 2010; Kuhlmann, Gómez-Rodríguez and Satta, 2011)

* Shi, Huang and Lee (2017, EMNLP)
I. UDpipe Pre-process

II. Feature Extraction

III. Unlabeled Parsing

IV. Arc Labeling

Compact (2) Feature Set

- Eisner’s
  - head
  - modifier

- Arc-eager
  - stack top
  - buffer top

- Arc-hybrid
  - stack top
  - buffer top

Scoring function: deep bi-affine
(Dozat and Manning, 2017)
<table>
<thead>
<tr>
<th>Ensembling</th>
<th>Full</th>
<th>Single Arc-eager</th>
<th>Single Arc-hybrid</th>
<th>Single Eisner’s</th>
</tr>
</thead>
<tbody>
<tr>
<td>LAS</td>
<td>75.00</td>
<td>74.32</td>
<td>74.00</td>
<td>73.75</td>
</tr>
<tr>
<td></td>
<td>75</td>
<td>74.5</td>
<td>74</td>
<td>74</td>
</tr>
<tr>
<td></td>
<td>74</td>
<td>73.5</td>
<td>73.75</td>
<td></td>
</tr>
<tr>
<td></td>
<td>73</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
I. UDPipe Pre-process

II. Feature Extraction

III. Unlabeled Parsing

IV. Arc Labeling

Multi-layer perceptron

concat( head, modifier )

nsubj | obj | .......

…….
I. UDPipe Pre-process

II. Feature Extraction

III. Unlabeled Parsing

IV. Arc Labeling

Effect of Ensemble

LAS

Full: 75.00

Single Labeler: 74.69
## Results — Official Ranking

<table>
<thead>
<tr>
<th>Category</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Big Treebanks</td>
<td>2</td>
</tr>
<tr>
<td>Small Treebanks</td>
<td>1</td>
</tr>
<tr>
<td>PUD Treebanks</td>
<td>2</td>
</tr>
<tr>
<td>Surprise Languages</td>
<td>1</td>
</tr>
<tr>
<td><strong>Overall</strong></td>
<td><strong>2</strong></td>
</tr>
</tbody>
</table>
Strategies — Small Treebanks

- Task finetune
  - fr model
    - Finetune on fr
      - All tasks
  - Task finetune
    - fr_partut model
      - Finetune on fr_partut
        - All tasks
  - Task finetune
    - fr_sequoia model
      - Finetune on fr_sequoia
        - All tasks

- Combined model

- Train on: \{fr, fr_partut, fr_sequoia\}
  - All tasks
Results — Small Treebanks

<table>
<thead>
<tr>
<th>Train Treebank</th>
<th>Test Treebank</th>
</tr>
</thead>
<tbody>
<tr>
<td>fr</td>
<td>fr_partut</td>
</tr>
<tr>
<td>fr</td>
<td>84.09</td>
</tr>
<tr>
<td>fr_partut</td>
<td></td>
</tr>
<tr>
<td>fr_sequoia</td>
<td></td>
</tr>
<tr>
<td>fr_sequoia</td>
<td>84.65</td>
</tr>
<tr>
<td>fr_sequoia</td>
<td></td>
</tr>
<tr>
<td>fr_sequoia</td>
<td>84.65</td>
</tr>
<tr>
<td>Combined</td>
<td>87.57</td>
</tr>
<tr>
<td>+Finetune</td>
<td>87.87</td>
</tr>
</tbody>
</table>

* UAS results on dev set, using gold segmentation
Strategies — Surprise Languages

- Train on a source language (selected via WALS)
- Delexicalized parser

Bi-directional LSTM

concat(
UPOS tag
Bag of Morphology tags
Max pooling
Morphology tags

parsing
## Results — Surprise Languages

<table>
<thead>
<tr>
<th>Target</th>
<th>Source*</th>
<th>Ranking</th>
</tr>
</thead>
<tbody>
<tr>
<td>Buryat</td>
<td>Hindi</td>
<td>2</td>
</tr>
<tr>
<td>Upper Sorbian</td>
<td>Czech</td>
<td>1</td>
</tr>
<tr>
<td>Kurmanji</td>
<td>Persian</td>
<td>1</td>
</tr>
<tr>
<td>North Sámi</td>
<td>Finnish</td>
<td>1</td>
</tr>
<tr>
<td><strong>Average</strong></td>
<td></td>
<td><strong>1</strong></td>
</tr>
</tbody>
</table>

*selected via WALS
Implementation

- Neural networks
- Parsing algorithms
- Hardware (×2)
- Training time: Approx. 1 week
Efficiency

Runtime (Hours) *

Stanford (Stanford) 16.27
C2L2 (Ithaca) 4.64
IMS (Stuttgart) 26.17
HIT-SCIR (Harbin) 8.88
LATTICE (Paris) 5.96

LAS
CPUs
RAM

76.30
4
16

75.00
2
8

74.42
12
64

72.11
1
8

70.93
8
32

* Not Benchmark Results
Combining Global Models for Parsing Universal Dependencies

argmax_{y \in Y} • Global transition-based models

• Ensemble

• Two-stage fine-tuning


Team C2L2 — Tianze Shi, Felix G. Wu, Xilun Chen, Yao Cheng