Arc-swift: A Novel Transition System for Dependency Parsing

Peng Qi and Christopher D. Manning
Stanford University
Dependency parsing

I ate fish with chopsticks.

http://corenlp.run, 2016/12/21
Dependency parsing

I ate fish with chopsticks.

I ate fish with ketchup.
Dependency parsing

Instrument?

Instrument!

http://corenlp.run, 2016/12/21
I ate fish with ketchup.
Transition systems

- arc-standard (Nivre, 2004)
- arc-eager (Nivre, 2003; Nivre 2008)
- arc-hybrid (Kuhlmann et al, 2011)
arc-eager

Shift

Left Arc

[root] ate fish

Right Arc

[root] I ate

Reduce

ate fish fork
Local transitions are difficult (arc-eager)

I ate fish with ketchup.
I ate fish with chopsticks.
Local transitions are difficult (arc-eager)

I ate fish with ketchup.
I ate fish with chopsticks.
Enhancing local transitions with features

<table>
<thead>
<tr>
<th>Single-word features (9)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$s_1.w; s_1.t; s_1.wt; s_2.w; s_2.t;$</td>
</tr>
<tr>
<td>$s_2.wt; b_1.w; b_1.t; b_1.wt$</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Word-pair features (8)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$s_1.wt \circ s_2.wt; s_1.wt \circ s_2.w; s_1.wts_2.t;$</td>
</tr>
<tr>
<td>$s_1.w \circ s_2.wt; s_1.t \circ s_2.wt; s_1.w \circ s_2.w$</td>
</tr>
<tr>
<td>$s_1.t \circ s_2.t; s_1.t \circ b_1.t$</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Three-word features (8)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$s_2.t \circ s_1.t \circ b_1.t; s_2.t \circ s_1.t \circ lc_1(s_1).t;$</td>
</tr>
<tr>
<td>$s_2.t \circ s_1.t \circ rc_1(s_1).t; s_2.t \circ s_1.t \circ lc_1(s_2).t;$</td>
</tr>
<tr>
<td>$s_2.t \circ s_1.t \circ rc_1(s_2).t; s_2.t \circ s_1.w \circ rc_1(s_2).t;$</td>
</tr>
<tr>
<td>$s_2.t \circ s_1.w \circ lc_1(s_1).t; s_2.t \circ s_1.w \circ b_1.t$</td>
</tr>
</tbody>
</table>

(Chen & Manning, 2014)
Local transitions are difficult (arc-eager)

I ate fish with ketchup.
I ate fish with chopsticks.

Key observation

Relatedness of “ate”, “fish”, and “*” determines which arc should be induced.
arc-swift

Shift

Left Arc [n]

Right Arc [n]
arc-swift

Shift

[left] ate fish

Left Arc [n]

ate Peng ’s fish

Right Arc [n]

ate fish fork

nmod:poss(Peng, fish)

nmod(ate, fork)
Resolving arc confusion with arc-swift
Model

“Head” and “dependent” representations

2-layer BiLSTM Parser

+ 32d POS embeddings

2-layer BiLSTM Tagger

100d GloVe word embeddings

See also: (Kiperwasser & Goldberg, 2016)
Model (cont’d)

See also: (Kiperwasser & Goldberg, 2016)
Model (cont’d)

Transition prediction

Softmax

[root] ate fish fork

See also: (Kiperwasser & Goldberg, 2016)
Model (cont’d)

See also: (Kiperwasser & Goldberg, 2016)
Model (cont’d)

Transition prediction

Softmax

s(ate $\rightarrow$ fork)  s(fish $\rightarrow$ fork)

H D H D H D H D

[root] ate fish fork

See also: (Kiperwasser & Goldberg, 2016)
Data & Evaluation

• Data
  • Penn Treebank (Marcus et al., 1999) WSJ portion converted to Stanford Dependencies (de Marneffe and Manning, 2008)
  • Universal Dependencies v1.3 (English) (Nivre et al., 2016)

• Evaluation
  • Unlabeled/labeled attachment score (UAS/LAS)
  • Punctuation removed
## Results

<table>
<thead>
<tr>
<th>Transition System</th>
<th>PTB UAS</th>
<th>PTB LAS</th>
<th>EN-UD UAS</th>
<th>EN-UD LAS</th>
</tr>
</thead>
<tbody>
<tr>
<td>arc-standard</td>
<td>94.0</td>
<td>91.7</td>
<td>85.6</td>
<td>81.5</td>
</tr>
<tr>
<td>arc-hybrid</td>
<td>94.0</td>
<td>91.8</td>
<td>85.4</td>
<td>81.4</td>
</tr>
<tr>
<td>arc-eager-S</td>
<td>93.8</td>
<td>91.7</td>
<td>85.2</td>
<td>81.2</td>
</tr>
<tr>
<td>arc-eager-R</td>
<td>93.9</td>
<td>91.7</td>
<td>85.4</td>
<td>81.3</td>
</tr>
<tr>
<td>arc-swift</td>
<td>94.3</td>
<td>92.2</td>
<td>86.1</td>
<td>82.2</td>
</tr>
</tbody>
</table>
## Results (cont’d)

<table>
<thead>
<tr>
<th>Implementation</th>
<th>TransSys</th>
<th>Notes</th>
<th>PTB UAS</th>
<th>PTB LAS</th>
<th>EN-UD UAS</th>
<th>EN-UD LAS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ours</td>
<td>arc-swift</td>
<td></td>
<td>94.3</td>
<td>92.2</td>
<td>86.1</td>
<td>82.2</td>
</tr>
<tr>
<td>Andor+ 2016</td>
<td>arc-standard</td>
<td>Feedforward, CRF loss, B=32</td>
<td>94.6</td>
<td>92.8</td>
<td>84.8*</td>
<td>80.4*</td>
</tr>
<tr>
<td>K&amp;G 2016</td>
<td>arc-hybrid</td>
<td>Dynamic oracle</td>
<td>93.6</td>
<td>91.5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Weiss+ 2015</td>
<td>arc-standard</td>
<td>B=8</td>
<td>94.0</td>
<td>92.1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C&amp;M 2014</td>
<td>arc-standard</td>
<td>Feedforward</td>
<td>91.8</td>
<td>89.6</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* https://github.com/tensorflow/models/blob/master/syntaxnet/g3doc/universal.md
Significance test (UAS/LAS)

<table>
<thead>
<tr>
<th>Row &gt; col?</th>
<th>arc-swift</th>
<th>arc-eager-S</th>
<th>arc-standard</th>
<th>arc-hybrid</th>
<th>arc-eager-R</th>
</tr>
</thead>
<tbody>
<tr>
<td>arc-swift</td>
<td><em><strong>/</strong></em></td>
<td><em><strong>/</strong></em></td>
<td><em><strong>/</strong></em></td>
<td><em><strong>/</strong></em></td>
<td><em><strong>/</strong></em></td>
</tr>
<tr>
<td>arc-eager-S</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>*/-</td>
<td></td>
</tr>
<tr>
<td>arc-standard</td>
<td>-</td>
<td>-</td>
<td>*/-</td>
<td></td>
<td></td>
</tr>
<tr>
<td>arc-hybrid</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* = p<0.05, ** = p<0.01, *** = p<0.001

In a 10-group Bonferroni-Holm test
Attachment error by dependency length

(Not all baseline transition systems shown)
Error Reduction in Linguistic Categories

![Graph showing error reduction in linguistic categories with categories: PP attachment, Noun/Adj confusion, Conj attachment, and Adv attachment. The graph compares arc-eager-S, arc-standard, and arc-swift.]
### Computational Efficiency

<table>
<thead>
<tr>
<th></th>
<th>arc-swift</th>
<th>arc-eager beam=2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of transitions evaluated per step</td>
<td>124%</td>
<td>400%</td>
</tr>
<tr>
<td>Average length of transition sequences</td>
<td>77.5%</td>
<td>100%</td>
</tr>
<tr>
<td>UAS</td>
<td>+0.3%</td>
<td>+0.2%</td>
</tr>
<tr>
<td>LAS</td>
<td>+0.5%</td>
<td>+0.3%</td>
</tr>
</tbody>
</table>
Thank you!

https://github.com/qipeng/arc-swift

Staying for CoNLL?
Checkout our Shared Task presentation!