Strongly Incremental Constituency Parsing with Graph Neural Networks

Kaiyu Yang and Jia Deng
Constituency Parsing

Arthur is King of the Britons

\[
\text{S} \\
\text{NP} \\
\text{Arthur} \\
\text{is} \\
\text{VP} \\
\text{NP} \\
\text{King} \\
\text{of} \\
\text{PP} \\
\text{NP} \\
\text{the} \\
\text{Britons}
\]
Shift-Reduce Parsers

Arthur is King of the Britons
Shift-Reduce Parsers

• SHIFT: Move the next word into the stack

Arthur is King of the Britons
Shift-Reduce Parsers

- SHIFT: Move the next word into the stack

shift, unary_reduce(NP)

Arthur is King of the Britons
Shift-Reduce Parsers

- SHIFT: Move the next word into the stack

shift, unary_reduce(NP)
Shift-Reduce Parsers

- SHIFT: Move the next word into the stack

Arthur is King of the Britons

shift, unary_reduce(NP), shift
Shift-Reduce Parsers

- **SHIFT**: Move the next word into the stack

Arthur is King of the Britons

shift, unary_reduce(NP), shift, shift, shift, binary_reduce(NP), shift, shift, shift, binary_reduce(PP), binary_reduce(NP), binary_reduce(VP), binary_reduce(S), King of the Britons
Arthur is King of the Britons

- SHIFT: Move the next word into the stack

\[\text{shift, unary\_reduce}(\text{NP}), \text{shift, shift, unary\_reduce}(\text{NP})\]
Shift-Reduce Parsers

• SHIFT: Move the next word into the stack

shift, unary_reduce(NP), shift, shift, unary_reduce(NP)
Shift-Reduce Parsers

- SHIFT: Move the next word into the stack

shift, unary_reduce(NP), shift, shift, unary_reduce(NP), shift

Arthur is King of the Britons
Shift-Reduce Parsers

- ** shift, unary_reduce(NP), shift, shift, unary_reduce(NP), shift, shift, unary_reduce(NP), shift, shift, unary_reduce(NP), shift, shift

  **SP**

  - **shift**: Move the next word into the stack

  - **SHIFT**: Move the next word into the stack
Shift-Reduce Parsers

- **SHIFT**: Move the next word into the stack
- **REDUCE**: Combine the top two elements in the stack

Arthur is King of the Britons

shift, unary_reduce(NP), shift, shift, unary_reduce(NP), shift, shift, **shift**
Shift-Reduce Parsers

- SHIFT: Move the next word into the stack
- REDUCE: Combine the top two elements in the stack

shift, unary_reduce(NP), shift, shift, unary_reduce(NP), shift, shift, shift, binary_reduce(NP)
Shift-Reduce Parsers

- SHIFT: Move the next word into the stack
- REDUCE: Combine the top two elements in the stack

shift, unaryReduce(NP), shift, shift, unaryReduce(NP), shift, shift, shift, binaryReduce(NP)
Shift-Reduce Parsers

- **SHIFT**: Move the next word into the stack
- **REDUCE**: Combine the top two elements in the stack

shift, unary_reduce(NP), shift, shift, unary_reduce(NP), shift, shift, shift, binary_reduce(NP), binary_reduce(PP)
Shift-Reduce Parsers

shift, unary_reduce(NP), shift, shift, unary_reduce(NP), shift, shift, shift, binary_reduce(NP), binary_reduce(PP)
Shift-Reduce Parsers

shift, unary_reduce(NP), shift, shift, unary_reduce(NP), shift, shift, shift, binary_reduce(NP), binary_reduce(PP), binary_reduce(NP)
Shift-Reduce Parsers

shift, unary_reduce(NP), shift, shift, unary_reduce(NP), shift, shift, shift, binary_reduce(NP),
binary_reduce(PP), binary_reduce(NP)
Shift-Reduce Parsers

shift, unary_reduce(NP), shift, shift, unary_reduce(NP), shift, shift, shift, binary_reduce(NP), binary_reduce(PP), binary_reduce(NP), binary_reduce(VP)
Arthur is the King of the Britons

shift, unary_reduce(NP), shift, shift, unary_reduce(NP), shift, shift, shift, binary_reduce(NP), binary_reduce(PP), binary_reduce(NP), binary_reduce(VP)
Shift-Reduce Parsers

shift, unary_reduce(NP), shift, shift, unary_reduce(NP), shift, shift, shift, binary_reduce(NP), binary_reduce(PP), binary_reduce(NP), binary_reduce(VP), binary_reduce(S)
Shift-Reduce Parsers

shift, unary_reduce(NP), shift, shift, unary_reduce(NP), shift, shift, shift, binary_reduce(NP),
binary_reduce(PP), binary_reduce(NP), binary_reduce(VP), binary_reduce(S)
Strongly Incremental Parsing

• Shift-reduce parsers differ from human parsing
• Human parsing appears to be strongly incremental

[Marslen-Wilson, 1973]
[Sturt and Lombardo, 2005]
[Stabler, 2015]
Strongly Incremental Parsing

\[ \text{Arthur is King of the Britons} \]

- Human parsing appears to be \textit{strongly incremental}:
  - One word per step: no more, no less
Strongly Incremental Parsing

• Human parsing appears to be *strongly incremental*:
  • One word per step: no more, no less
Strongly Incremental Parsing

• Human parsing appears to be *strongly incremental*:
  • One word per step: no more, no less
Strongly Incremental Parsing

Arthur is King of the Britons

- Human parsing appears to be *strongly incremental*:
  - One word per step: no more, no less
Strongly Incremental Parsing

- Human parsing appears to be *strongly incremental*:
  - One word per step: no more, no less
Strongly Incremental Parsing

- Human parsing appears to be *strongly incremental*:
  - One word per step: no more, no less

```
S
  NP
    Arthur
  VP
    is
  NP
    King
```

*of the Britons*
Strongly Incremental Parsing

- Human parsing appears to be *strongly incremental*:
  - One word per step: no more, no less
  - A single connected partial parse tree
Attach-Juxtapose Transition System

- We propose a strongly incremental transition system named *attach-juxtapose*
Attach-Juxtapose Transition System

• We propose a strongly incremental transition system named *attach-juxtapose*
• The state is a partial tree and the next word
• Actions determine *where* and *how* to integrate the next word
Where to Add the New Word?

• The new word is to the right of existing words, so it must appear on the *rightmost chain*

```
(S
  (NP Arthur)
  (VP is
    (NP King)
  )
)
```
Where to Add the New Word?

• The new word is to the right of existing words, so it must appear on the *rightmost chain*
How to Add the New Word?

```
S
/   |
NP   VP
|    |
Arthur is NP
    |
King of
```
How to Add the New Word?

S
 NP Arthur
 VP is NP King of
How to Add the New Word?

Arthur is VP NP King of

S

NP VP

Arthur is NP

King

of
How to Add the New Word?
How to Add the New Word?

ATTACH
How to Add the New Word?

Attaching the new word "of" to the NP "King".
How to Add the New Word?
How to Add the New Word?

ATTACH

JUXTAPOSE

S

NP

Arthur

is

VP

King

S

NP

Arthur

is

VP

King

of

of
Action Generation with Graph Neural Networks

- Encoder: BERT/XLNet + additional self-attention layers
Action Generation with Graph Neural Networks

- Encoder: BERT/XLNet + additional self-attention layers
- Decoder: Generate attach-juxtapose actions by applying GNNs on the partial tree
Action Generation with Graph Neural Networks

- Encoder: BERT/XLNet + additional self-attention layers
- Decoder: Generate attach-juxtapose actions by applying GNNs on the partial tree
Experimental Results

- Competitive with state of the art on Penn Treebank

<table>
<thead>
<tr>
<th>Model</th>
<th>EM</th>
<th>F1</th>
<th>LP</th>
<th>LR</th>
<th>#Params</th>
</tr>
</thead>
<tbody>
<tr>
<td>Liu and Zhang [22]</td>
<td>-</td>
<td>91.8</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Liu and Zhang [22] (BERT) †</td>
<td>57.05</td>
<td>95.71</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Kitaev and Klein [21]</td>
<td>47.31</td>
<td>93.55</td>
<td>93.90</td>
<td>93.20</td>
<td>26M</td>
</tr>
<tr>
<td>Kitaev and Klein [21] (ELMo)</td>
<td>53.06</td>
<td>95.13</td>
<td>95.40</td>
<td>94.85</td>
<td>107M</td>
</tr>
<tr>
<td>Kitaev et al. [20] (BERT)</td>
<td>-</td>
<td>95.59</td>
<td>95.46</td>
<td>95.73</td>
<td>342M</td>
</tr>
<tr>
<td>Zhou and Zhao [49] (GloVe) *</td>
<td>47.72</td>
<td>93.78</td>
<td>93.92</td>
<td>93.64</td>
<td>51M</td>
</tr>
<tr>
<td>Zhou and Zhao [49] (BERT) *</td>
<td>55.84</td>
<td>95.84</td>
<td>95.98</td>
<td>95.70</td>
<td>349M</td>
</tr>
<tr>
<td>Zhou and Zhao [49] (XLNet) *</td>
<td>58.73</td>
<td>96.33</td>
<td>96.46</td>
<td>96.21</td>
<td>374M</td>
</tr>
<tr>
<td>Mrini et al. [27] (XLNet) *</td>
<td>58.65</td>
<td><strong>96.38</strong></td>
<td>96.53</td>
<td><strong>96.24</strong></td>
<td>459M</td>
</tr>
<tr>
<td>Ours (BERT)</td>
<td>57.29 ± 0.57</td>
<td>95.79 ± 0.05</td>
<td>96.04 ± 0.05</td>
<td>95.55 ± 0.06</td>
<td>377M</td>
</tr>
<tr>
<td>Ours (XLNet)</td>
<td><strong>59.17</strong> ± 0.33</td>
<td><strong>96.34</strong> ± 0.03</td>
<td><strong>96.55</strong> ± 0.02</td>
<td>96.13 ± 0.04</td>
<td>391M</td>
</tr>
</tbody>
</table>
Experimental Results

- Competitive with state of the art on Penn Treebank
- Improves upon state of the art on Chinese Treebank

<table>
<thead>
<tr>
<th>Model</th>
<th>EM</th>
<th>F1</th>
<th>LP</th>
<th>LR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kitaev et al. [20]</td>
<td>-</td>
<td>91.75</td>
<td>91.96</td>
<td>91.55</td>
</tr>
<tr>
<td>Kitaev et al. [20] (BERT) †</td>
<td>44.42</td>
<td>92.14</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Zhou and Zhao [49] *</td>
<td>-</td>
<td>92.18</td>
<td>92.33</td>
<td>92.03</td>
</tr>
<tr>
<td>Mrini et al. [27] (BERT) *</td>
<td>-</td>
<td>92.64</td>
<td>93.45</td>
<td>91.85</td>
</tr>
<tr>
<td>Liu and Zhang [22]</td>
<td>-</td>
<td>86.1</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Liu and Zhang [22] (BERT) †</td>
<td>44.94</td>
<td>91.81</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Ours (BERT)</td>
<td>49.72±0.83</td>
<td>93.59±0.26</td>
<td>93.80±0.26</td>
<td>93.40±0.28</td>
</tr>
</tbody>
</table>
Strongly Incremental Constituency Parsing with Graph Neural Networks

Kaiyu Yang and Jia Deng

https://github.com/princeton-vl/attach-juxtapose-parser