Global Neural CCG Parsing with Optimality Guarantees

Kenton Lee       Mike Lewis†       Luke Zettlemoyer

University of Washington

† Now at Facebook AI Research
This Talk

Challenge:
Global models (e.g. Recursive NNs) break dynamic programs
This Talk

Challenge:
Global models (e.g. Recursive NNs) break dynamic programs

Our approach:
Combine local and global models in A* parser

Result:
Global model with exact inference
Parsing with Hypergraphs

Fruit flies like bananas

Klein and Manning, 2001
Fruit flies like bananas

\[
\begin{align*}
\text{NP/NP} & \rightarrow \text{NP} \\
\text{NP} & \rightarrow (S\backslash NP)/NP \\
(S\backslash NP)/NP & \rightarrow \text{NP} \\
(S\backslash NP)/NP & \rightarrow S \\
S & \rightarrow < \\
\end{align*}
\]
Parsing with Hypergraphs

Fruit flies like bananas

Nodes represent partial parses

Klein and Manning, 2001
Parsing with Hypergraphs

Fruit flies like bananas

Hyperedges represent rule productions

Klein and Manning, 2001
Fruit flies like bananas

Path \( y = \{e_1, \ldots, e_m\} \) represents a parse derivation
Parsing with Hypergraphs

Input

Fruit flies like bananas

Output

```
NP/NP  NP  (S\NP)/NP  NP  NP
      NP  S\NP     
        S  

NP/NP  NP  (S\NP)/NP  NP  NP
      NP  S\NP     
        S  

NP/NP  NP  (S\NP)/NP  NP  NP
      NP  S\NP     
        S  
```

Fruit flies like bananas

0
Parsing with Hypergraphs

Fruit flies like bananas
Fruit flies like bananas

Each hyperedge $e$ is weighted with a score $g(e)$.
Fruit flies like bananas

Score of parse derivation:

$$g(y) = \sum_{e \in y} g(e)$$
Fruit flies like bananas.
Parsing with Hypergraphs
Parsing with Hypergraphs

- Predicted parse: \( y^* = \arg\max_{y \in Y} g(y) \)
- Exponential number of nodes

\[ \rightarrow \quad \text{Intractable inference} \]
Managing Intractable Search Spaces

Approximate inference with global expressivity, e.g.

- Greedy / beam search:
  - Nivre, 2008
  - Chen and Manning, 2014
  - Andor et al., 2016

- Reranking:
  - Charniak and Johnson, 2005
  - Huang, 2008
  - Socher et al., 2013
Locally Factored Parsing

Scores condition on local structures

- Make locality assumptions:
  - e.g. features are local to CFG productions
- Polynomial number of nodes
- Dynamic programs enable tractable inference
Locally Factored Parsing

Scores condition on local structures

Dynamic programs with locally factored models, e.g.

- CKY:
  - Collins, 1997
  - Durrett and Klein, 2015

- Minimum spanning tree:
  - McDonald et al., 2005
  - Kiperwasser and Goldberg, 2016
Locally Factored Parsing

Scores condition on local structures

Dynamic programs with locally factored models, e.g.

Recursive neural networks break dynamic programs!

- Minimum spanning tree:
  - McDonald et al., 2005
  - Kiperwasser and Goldberg, 2016
Local vs. Global Models

Local model:

\[ y^* = \arg\max_{y \in Y} (g_{\text{local}}(y)) \]

Efficient  Inexpressive

Global model:

\[ y^* = \arg\max_{y \in Y} (g_{\text{global}}(y)) \]

Intractable  Expressive
This Work

Combined model:

\[ y^* = \text{argmax}_{y \in Y} \left( g_{\text{local}}(y) + g_{\text{global}}(y) \right) \]

Efficient

Expressive
Outline

- Background: A* parsing
- Combined global and local parsing model
- Learning to search accurately and efficiently
- Experiments on CCGBank
A* Parsing

\[ y^* = \arg \max_{y \in Y} g(y) \]

- Search in the space of partial parses
- First explored full parse guaranteed to be optimal
Fruit flies like bananas

\[
\begin{align*}
S & \rightarrow NP \\
(S \setminus NP) & \rightarrow NP \\
NP & \rightarrow NP
\end{align*}
\]

Partial parse
A* Parsing

Fruit flies like bananas

Partial parse
A* Parsing

\[ f\left( \frac{\text{Fruit flies}}{?} \right) \frac{\text{like}}{(S\setminus NP)/NP} \frac{\text{bananas}}{NP} \]

Exploration priority

Partial parse
A* Parsing

Exploration priority

\[ f(\text{Fruit flies like bananas}) = g(\text{Fruit flies like bananas}) + h(\text{Fruit flies like bananas}) \]

Inside score

Admissible A* heuristic

Fruit flies like bananas

\[ \frac{\text{Fruit flies}}{\text{S\textbackslash NP}} \quad \frac{\text{like}}{\text{S\textbackslash NP}} \quad \frac{\text{bananas}}{\text{NP}} \]

\[ \frac{(S\textbackslash NP)/NP}{\text{NP}} \quad \frac{\text{NP}}{\text{S\textbackslash NP}} \]
A* Parsing

![Diagram of A* Parsing]
### A* Parsing

```
Fruit flies like bananas
NP    NP\NP   (S\NP)/NP   NP
       NP\NP   (S\NP)/NP   S\NP
       S\NP   S\NP
       S
```

#### Agenda position $f(y)$ $y$

<table>
<thead>
<tr>
<th>Agenda position</th>
<th>$f(y)$</th>
<th>y</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>4.5</td>
<td>bananas</td>
</tr>
<tr>
<td>2</td>
<td>3.1</td>
<td>like</td>
</tr>
<tr>
<td>3</td>
<td>1.9</td>
<td>Fruit</td>
</tr>
<tr>
<td>4</td>
<td>-0.5</td>
<td>Fruit</td>
</tr>
</tbody>
</table>
### A* Parsing

<table>
<thead>
<tr>
<th>Agenda position</th>
<th>$f(y)$</th>
<th>$y$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>4.5</td>
<td>bananas $NP$</td>
</tr>
<tr>
<td>2</td>
<td>3.1</td>
<td>like $(S\backslash NP)/NP$</td>
</tr>
<tr>
<td>3</td>
<td>1.9</td>
<td>Fruit $NP$</td>
</tr>
<tr>
<td>4</td>
<td>-0.5</td>
<td>Fruit $NP/ NP$</td>
</tr>
</tbody>
</table>

**Diagram:**
- The diagram illustrates the parsing process with nodes representing different grammatical structures such as $NP$, $S\backslash NP$, and $S\backslash S$.
- The edges connect nodes indicating the progression of the parsing algorithm.
- The agenda and explored/unexplored states are marked to show the order of processing.

**Table:**
- The table shows the agenda position, the cost function $f(y)$, and the generated string $y$ for each step of the parsing algorithm.
A* Parsing

<table>
<thead>
<tr>
<th>Agenda position</th>
<th>$f(y)$</th>
<th>$y$</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>3.1</td>
<td>like $\frac{(S\backslash NP)/NP}{NP}$</td>
</tr>
<tr>
<td>3</td>
<td>1.9</td>
<td>Fruit $NP$</td>
</tr>
<tr>
<td>4</td>
<td>-0.5</td>
<td>Fruit $\frac{NP/\overline{NP}}{NP}$</td>
</tr>
</tbody>
</table>
### A* Parsing

![Diagram of A* Parsing]

- **Agenda position**
- **$f(y)$**
- **$y$**

<table>
<thead>
<tr>
<th>Agenda position</th>
<th>$f(y)$</th>
<th>$y$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>3.1</td>
<td>like $(S\backslash NP)/NP$</td>
</tr>
<tr>
<td>2</td>
<td>1.9</td>
<td>Fruit $NP$</td>
</tr>
<tr>
<td>3</td>
<td>-0.5</td>
<td>Fruit $NP/SP$</td>
</tr>
<tr>
<td>4</td>
<td>-1.3</td>
<td>flies $NP$</td>
</tr>
</tbody>
</table>
A* Parsing

Agenda position | $f(y)$  | $y$
--- | --- | ---
1 | 3.1 | $\frac{(S\backslash NP)/NP}{NP}$
2 | 1.9 | $\frac{NP}{NP}$
3 | -0.5 | $\frac{Fruit}{NP}$
4 | -1.3 | $\frac{flies}{NP}$
A* Parsing

<table>
<thead>
<tr>
<th>Agenda position</th>
<th>$f(y)$</th>
<th>$y$</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>1.9</td>
<td>Fruit $NP$</td>
</tr>
<tr>
<td>3</td>
<td>-0.5</td>
<td>Fruit $NP/NP$</td>
</tr>
<tr>
<td>4</td>
<td>-1.3</td>
<td>flies $NP$</td>
</tr>
</tbody>
</table>
A* Parsing
A* Parsing

<table>
<thead>
<tr>
<th>Agenda position</th>
<th>$f(y)$</th>
<th>$y$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2.1</td>
<td>like</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(NP/NP)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>S/NP</td>
</tr>
<tr>
<td>2</td>
<td>1.9</td>
<td>Fruit</td>
</tr>
<tr>
<td></td>
<td></td>
<td>NP</td>
</tr>
<tr>
<td>3</td>
<td>-0.5</td>
<td>Fruit</td>
</tr>
<tr>
<td></td>
<td></td>
<td>NP/NP</td>
</tr>
<tr>
<td>4</td>
<td>-1.3</td>
<td>flies</td>
</tr>
<tr>
<td></td>
<td></td>
<td>NP</td>
</tr>
</tbody>
</table>

Like bananas (NP/S)
A* Parsing

<table>
<thead>
<tr>
<th>Agenda position</th>
<th>$f(y)$</th>
<th>$y$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1.9</td>
<td>Fruit $NP$</td>
</tr>
<tr>
<td>2</td>
<td>-1.5</td>
<td>like $(S\backslash S)/NP$</td>
</tr>
<tr>
<td>3</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>4</td>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>
Locally Factored Model

Supertag-factored A* CCG Parser (Lewis et al, 2016):

\[
\begin{align*}
\text{Fruit} & \quad \text{flies} & \quad \text{like} & \quad \text{bananas} \\
\frac{NP/NP}{NP} & \quad \frac{NP}{NP} & \quad \frac{(S\backslash NP)/NP}{S\backslash NP} & \quad \frac{NP}{NP} \\
\rightarrow & & \rightarrow & < \\
NP & & S\backslash NP & S
\end{align*}
\]
Locally Factored Model

Supertag-factored A* CCG Parser (Lewis et al, 2016):

\[ g_{local}(\text{Fruit NP/NP}) + g(\text{flies NP}) + g(\text{like (S\ NP)/NP}) + g(\text{bananas NP}) \]
Locally Factored Model

Supertag-factored $A^*$ CCG Parser (Lewis et al, 2016):

\[
\begin{array}{ccccccc}
\text{Fruit} & \text{flies} & \text{like} & \text{bananas} \\
? & (S\backslash NP)/NP & NP & S\backslash NP \\
\end{array}
\]
Locally Factored Model

Supertag-factored A* CCG Parser (Lewis et al, 2016):

\[ g_{local}(\text{Fruit flies ? like bananas}) = g\left(\frac{(S\backslash NP)/NP}{NP}\right) + g\left(\frac{\text{bananas}}{NP}\right) \]
Locally Factored Model

Supertag-factored A* CCG Parser (Lewis et al, 2016):

\[ g_{local}(\text{?}) : g\left(\frac{(S\backslash NP)/NP}{NP}\right) + g\left(\frac{bananas}{NP}\right) \]

\[ h_{local}(\text{?}) : \max_{\text{tag}} g\left(\frac{\text{Fruit}}{\text{tag}}\right) + \max_{\text{tag}} g\left(\frac{\text{Flies}}{\text{tag}}\right) \]
Outline

- Background: A* parsing
- Combined global and local parsing model
- Learning to search accurately and efficiently
- Experiments on CCGBank
Global A* Parsing

\[ y^* = \arg \max_{y \in Y} g(y) \]

- First explored full parse guaranteed to be optimal
- Global search graph is exponential in sentence length
- Open question: Can we still learn to search efficiently?
Modeling Global Structure

\[ g_{\text{global}}(y) : \]

\[ h_{\text{global}}(y) : \]
Modeling Global Structure

\[ g(y) = g_{\text{global}}(y) \]

\[ h(y) = 0 \]
Modeling Global Structure

\[ g(y) = g_{\text{global}}(y) \]

\[ h(y) = 0 \]
Any locally factored model with an admissible A* heuristic

Non-positive global model
Division of Labor

\[ g(y) = g_{local}(y) + g_{global}(y) \]

- Limited expressivity
- Provides guidance with an A* heuristic

- Global expressivity
- Discriminative only when necessary
Global Model: $g_{global}(y)$
Non-positive Global Model

\[ g_{\text{global}}(\cdot) = \log(\sigma(w \cdot \cdot)) \]
Division of Labor

\[ g(y) = g_{local}(y) + g_{global}(y) \]

- Limited expressivity
- Provides guidance with an A* heuristic

- Global expressivity
- Discriminative only when necessary
Outline

- Background: A* parsing
- Combined global and local parsing model
- Learning to search accurately and efficiently
- Experiments on CCGBank
Learning with A*
Learning with A*

Agenda position | $f(y)$ | $y$ | Is correct?
--- | --- | --- | ---
1 | 4.5 | bananas $\frac{NP}{NP}$ | ✓
2 | 3.1 | like $\frac{(S\backslash NP)/NP}{NP}$ | ✓
3 | 1.9 | Fruit $\frac{NP}{NP}$ | ✗
4 | -0.5 | Fruit $\frac{NP/NP}{NP}$ | ✓
Learning with A*

<table>
<thead>
<tr>
<th>Agenda position</th>
<th>$f(y)$</th>
<th>$y$</th>
<th>Is correct?</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1.9</td>
<td>Fruit $NP$</td>
<td>✗</td>
</tr>
<tr>
<td>2</td>
<td>-0.5</td>
<td>Fruit $NP/NP$</td>
<td>✗</td>
</tr>
<tr>
<td>3</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>4</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>

Fruit flies like bananas
$NP \rightarrow NP/NP$ like bananas
$NP/NP \rightarrow (S/NP)/NP$ bananas
$NP \rightarrow S/\NP$ bananas
$NP/\NP \rightarrow (S/\NP)/NP$ bananas
$NP \rightarrow S$ bananas
Learning with A*

### Agenda violation:
incorrect partial parse explored

<table>
<thead>
<tr>
<th>Agenda position</th>
<th>$f(y)$</th>
<th>$y$</th>
<th>Is correct?</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1.9</td>
<td>Fruit $\quad NP$</td>
<td>✗</td>
</tr>
<tr>
<td>2</td>
<td>-0.5</td>
<td>Fruit $\quad NP/NP$</td>
<td>✓</td>
</tr>
<tr>
<td>3</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>4</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>
Violation-based Loss

A: [ ... ]

<table>
<thead>
<tr>
<th>Agent position</th>
<th>$f(y)$</th>
<th>$y$</th>
<th>Is correct?</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>4.5</td>
<td>$N^P$</td>
<td>✓</td>
</tr>
<tr>
<td>2</td>
<td>3.1</td>
<td>$N^P$</td>
<td>✓</td>
</tr>
<tr>
<td>3</td>
<td>1.9</td>
<td>$N^P$</td>
<td>✓</td>
</tr>
<tr>
<td>4</td>
<td>-0.5</td>
<td>$N^P$</td>
<td>✓</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Agent position</th>
<th>$f(y)$</th>
<th>$y$</th>
<th>Is correct?</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1.9</td>
<td>$P^N$</td>
<td>✗</td>
</tr>
<tr>
<td>2</td>
<td>-0.5</td>
<td>$P^N$</td>
<td>✓</td>
</tr>
<tr>
<td>3</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>4</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>
Violation-based Loss

\[ L(A) = \sum_{t=1}^{T} \max_{y \in A_t} f(y) - \max_{y \in \text{GOLD}(A_t)} f(y) \]

**Top of agenda**

**Best gold partial parse**
Jointly Optimizing Accuracy and Efficiency

Correct partial parse can still be predicted via backtracking

<table>
<thead>
<tr>
<th>Agenda position</th>
<th>$f(y)$</th>
<th>$y$</th>
<th>Is correct?</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1.9</td>
<td>Fruit $\frac{NP}{NP}$</td>
<td>❌</td>
</tr>
<tr>
<td>2</td>
<td>-0.5</td>
<td>Fruit $\frac{NP/\overline{NP}}{NP}$</td>
<td>✔</td>
</tr>
<tr>
<td>3</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>4</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>
Jointly Optimizing Accuracy and Efficiency

Explicitly optimize for search efficiency!

<table>
<thead>
<tr>
<th>Agenda position</th>
<th>$f(y)$</th>
<th>$y$</th>
<th>Is correct?</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>…</td>
<td>…</td>
<td>…</td>
</tr>
<tr>
<td>4</td>
<td>…</td>
<td>…</td>
<td>…</td>
</tr>
</tbody>
</table>
Outline

❖ Background: A* parsing
❖ Combined global and local parsing model
❖ Learning to search accurately and efficiently
❖ Experiments on CCGBank
Experimental Setup

- $g_{local}(y)$: supertag-factored model from Lewis et al. (2016)
- Evaluate on CCGBank (Hockenmaier & Steedman, 2007)
- Comparisons:

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Is global?</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Is exact?</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Experimental Setup

- $\text{Global}(y)$: supertag-factored model from Lewis et al. (2016)
- Evaluate on CCGBank (Hockenmaier & Steedman, 2007)
- Comparisons:

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Is global?</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Is exact?</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
</tbody>
</table>
## CCG Parsing Results

### Test F1 (%)

<table>
<thead>
<tr>
<th></th>
<th>84.0</th>
<th>85.0</th>
<th>86.0</th>
<th>87.0</th>
<th>88.0</th>
<th>89.0</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clark &amp; Curran (2007)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>85.2</td>
<td></td>
</tr>
<tr>
<td>Xu et al. (2015)</td>
<td></td>
<td></td>
<td></td>
<td>87.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lewis et al. (2016)</td>
<td></td>
<td></td>
<td>88.1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Vaswani et al. (2016)</td>
<td></td>
<td></td>
<td></td>
<td>88.3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Global A*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>88.7</td>
</tr>
</tbody>
</table>

### Is global?

|                | ✓    | ✓    | ✓    | ✓    | ✓    |

### Is exact?

|                | ✓    | ✓    | ✓    | ✓    |
CCG Parsing Results

- Optimal parse found for 99.9% of sentences
- Explores only 190 partial parses on average

<table>
<thead>
<tr>
<th>Is global?</th>
<th>✓</th>
<th>✓</th>
<th>✓</th>
</tr>
</thead>
<tbody>
<tr>
<td>Is exact?</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
</tbody>
</table>
Decoder Comparisons

- Development F1 (%)
- Speed (sentences / second)

10-best Reranking: 87.9, 3.2
100-best Reranking: 88.2, 0.4
4-best Beam Search: 88.3, 4.0
Global A*: 88.4, 27.1
Context Ablation

- Development F1 (%)
- Number of explorations (lower is better)

Global A*
- 88.4
- 309.6

Global A* without context
- 88.1
- 610.5

Example sentence:
Fruit flies like (S\NP) bananas.
The favorite U.S. small business is one whose research and development can be milked for future Japanese use.
Conclusion

❖ Combining local and global models enables exact inference with global features

❖ Efficient decoding by learning to search

❖ State of the art for CCG parsing

❖ Applicable to other structured prediction tasks