Interactive Learning of Parsers from Weak Supervision

Luke Zettlemoyer

with Luheng He, Kenton Lee, Mike Lewis, Julian Michael
Interpreting Language

Sentence

Semantic Parser

Meaning Representation

Executor

Response
Semantic Parsing: QA

How many people live in Seattle?

Semantic Parser

SELECT Population FROM CityData where City=="Seattle";

Executor

620,778

[Wong & Mooney 2007],
[Zettlemoyer & Collins 2005, 2007],
[Kwiatkowski et.al 2010, 2011],
[Liang et.al. 2011], [Cai & Yates 2013],
[Berant et.al. 2013,2014,2015],
[Kwiatkowski et.al. 2013],
[Reddy et.al, 2014,2016]
Go to the third junction and take a left

\[
\text{(do-seq})(\text{do-n-times} 3
\quad (\text{move-to forward-loc}
\quad (\text{do-until}
\quad (\text{junction current-loc}
\quad (\text{move-to forward-loc}))))
\quad (\text{turn-right})\]

Semantic Parsing: Instructions

Chen & Mooney 2011
Matuszek et.al. 2012
Artzi & Zettlemoyer 2013
Mei et.al. 2015
Somerset Maugham was a British playwright, novelist and short story writer.

Knowledge Base (KB)

Semantic Parser

<table>
<thead>
<tr>
<th>S. Maugham</th>
<th>Nationality</th>
<th>United Kingdom</th>
</tr>
</thead>
<tbody>
<tr>
<td>S. Maugham</td>
<td>Profession</td>
<td>Novelist</td>
</tr>
</tbody>
</table>

[Krishnamurthy and Mitchell; 2012, 2014][Choi et al., 2015]
Semantic Parsing: Complex Structure

How many people live in Seattle

Latent

620,778
Lots of Different Applications

We are doing semantic analysis for:

- Visual Semantic Role Labeling [Yatskar et al, 2016]
- Visual Question Answering [FitzGerald et al, in prep]
- Language to Code [Lin et al, in prep]
- Entity-entity sentiment [Choi et al, 2016]
- Understanding Cooking Recipes [Kiddon et al, 2016]
- Zero-shot Relation Extraction [Levy et al, in review]
- Interactive Learning for NLIDBs [Iyer, et al, in review]

Challenge: typically gather data and learn model from scratch in each case…
Understanding Cooking Recipes

Amish Meatloaf (http://allrecipes.com/recipe/amish-meatloaf/, recipe condensed)

Ingredients
2 pounds ground beef
2 1/2 cups crushed butter-flavored crackers
1 small onion, chopped
2 eggs
3/4 cup ketchup
1/4 cup brown sugar
2 slices bacon

Preheat the oven to 350 degrees F (175 degrees C).
In a medium bowl, mix together ground beef, crushed crackers, onion, eggs, ketchup, and brown sugar until well blended.
Press into a 9x5 inch loaf pan.
Lay the two slices of bacon over the top.
Bake for 1 hour, or until cooked through.

Approach: unsupervised learning for actions and object flow

Open Question:

• Can we build an off-the-shelf parser that would help here?

[Kiddon et al 2015, 2016]
Towards Broad Coverage Semantic Parsing

• Can we crowdsource semantics?
• Train with latent syntax?
• Build fast and accurate parsers?
• Actively select which data to label?
Semantic Role Labeling (SRL)

who did what to whom, when and where?

- They 
- increased 
- the rent 
- drastically 
- this year

- Agent
- Predicate
- Patent
- Manner
- Time

- Defining a set of roles can be difficult
- Existing formulations have used different sets
Existing SRL Formulations and Their Frame Inventories

**FrameNet**
1000+ semantic frames, roles (frame elements) shared across frames

**PropBank**
10,000+ frame files with predicate-specific roles

**Frame: Change_position_on_a_scale**
This frame consists of words that indicate the change of an Item's position on a scale (the Attribute) from a starting point (Initial_value) to an end point (Final_value). The direction (Path) …

**Lexical Units:**
..., reach.v, rise.n, rise.v, rocket.v, shift.n, …

**Roleset Id:** rise.01, go up

- **Arg1-:** Logical subject, patient, thing rising
- **Arg2-EXT:** EXT, amount risen
- **Arg3-DIR:** start point
- **Arg4-LOC:** end point
- **Argm-LOC:** medium

Unified Verb Index, University of Colorado [http://verbs.colorado.edu/verb-index/](http://verbs.colorado.edu/verb-index/)
PropBank Annotation Guidelines, Bonial et al., 2010
FrameNet II: Extended theory and practice, Ruppenhofer et al., 2006
Our Annotation Scheme

Given sentence and a verb:

They *increased* the rent this year.

Step 1: Ask a question about the verb:

Who increased something?

Step 2: Answer with words in the sentence:

They

Step 3: Repeat, write as many QA pairs as possible...

What is increased?

the rent

When is something increased?

this year

[He et al 2015]
Our Method: Q/A Pairs for Semantic Relations

The rent rose 10% from $3000 to $3300.

**Wh-Question**

What rose?

How much did something rise?

What did something rise from?

What did something rise to?

**Answer**

the rent

10%

$3000

$3300
Dataset Statistics

- Sentences: 1,241 (newswire) vs 1,959 (Wikipedia)
- Verbs: 3,336 (newswire) vs 4,440 (Wikipedia)
- QA Pairs: 8,109 (newswire) vs 10,798 (Wikipedia)
Cost and Speed

- Part-time freelancers from upwork.com (hourly rate: $10)
- ~2h screening process for native English proficiency
## Wh-words vs. PropBank Roles

<table>
<thead>
<tr>
<th></th>
<th>Who</th>
<th>What</th>
<th>When</th>
<th>Where</th>
<th>Why</th>
<th>How</th>
<th>HowMuch</th>
</tr>
</thead>
<tbody>
<tr>
<td>ARG0</td>
<td>1575</td>
<td>414</td>
<td>3</td>
<td>5</td>
<td>17</td>
<td>28</td>
<td>2</td>
</tr>
<tr>
<td>ARG1</td>
<td>285</td>
<td>2481</td>
<td>4</td>
<td>25</td>
<td>20</td>
<td>23</td>
<td>95</td>
</tr>
<tr>
<td>ARG2</td>
<td>85</td>
<td>364</td>
<td>2</td>
<td>49</td>
<td>17</td>
<td>51</td>
<td>74</td>
</tr>
<tr>
<td>ARG3</td>
<td>11</td>
<td>62</td>
<td>7</td>
<td>8</td>
<td>4</td>
<td>16</td>
<td>31</td>
</tr>
<tr>
<td>ARG4</td>
<td>2</td>
<td>30</td>
<td>5</td>
<td>11</td>
<td>2</td>
<td>4</td>
<td>30</td>
</tr>
<tr>
<td>ARG5</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>AM-ADV</td>
<td>5</td>
<td>44</td>
<td>9</td>
<td>2</td>
<td>25</td>
<td>27</td>
<td>6</td>
</tr>
<tr>
<td>AM-CAU</td>
<td>0</td>
<td>3</td>
<td>1</td>
<td>0</td>
<td>23</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>AM-DIR</td>
<td>0</td>
<td>6</td>
<td>1</td>
<td>13</td>
<td>0</td>
<td>4</td>
<td>0</td>
</tr>
<tr>
<td>AM-EXT</td>
<td>0</td>
<td>4</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>AM-LOC</td>
<td>1</td>
<td>35</td>
<td>10</td>
<td>89</td>
<td>0</td>
<td>13</td>
<td>11</td>
</tr>
<tr>
<td>AM-MNR</td>
<td>5</td>
<td>47</td>
<td>2</td>
<td>8</td>
<td>4</td>
<td>108</td>
<td>14</td>
</tr>
<tr>
<td>AM-PNC</td>
<td>2</td>
<td>21</td>
<td>0</td>
<td>1</td>
<td>39</td>
<td>7</td>
<td>2</td>
</tr>
<tr>
<td>AM-PRD</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>AM-TMP</td>
<td>2</td>
<td>51</td>
<td>341</td>
<td>2</td>
<td>11</td>
<td>20</td>
<td>10</td>
</tr>
</tbody>
</table>
Advantages
• Easily explained
• No pre-defined roles, few syntactic assumptions
• Can capture implicit arguments
• Generalizable across domains

Limitations
• Only modeling verbs (for now)
• Not annotating verb senses directly
• Can have multiple equivalent questions

Challenges
• What questions to ask?
• Quality - Can we get good Q/A pairs?
• Coverage - Can we get all the Q/A pairs?
Towards Broad Coverage Semantic Parsing

- Can we crowdsourcing semantics?
- Train with latent syntax?
- Build fast and accurate parsers?
- Actively select which data to label?
John denied the report

John refused to deny the report

John refused to confirm or deny the report
Joint vs. Pipelines

F1
CCG Dependencies

Include nearly all SRL dependencies:

\[
\begin{align*}
\text{John} & \quad \text{wanted} \quad \text{to confirm} \quad \text{the report} \\
\text{NP}_{\text{John}} & \quad (S\backslash\text{NP}_x)/(S\backslash\text{NP}_x) & \quad (S\backslash\text{NP}_x)/\text{NP}_y & \quad \text{确认} \rightarrow x, \text{确认} \rightarrow y \\
\text{wanted} & \rightarrow x & \text{to confirm} & \rightarrow x, \text{确认} \rightarrow y \\
\text{the report} & \rightarrow x & \text{确认} & \rightarrow \text{report}, \text{确认} \rightarrow x \\
\text{wanted} & \rightarrow x & \text{确认} & \rightarrow \text{report}, \text{确认} \rightarrow x, \text{wanted} \rightarrow x \\
\text{S} & \quad \text{确认} \rightarrow \text{report}, \text{确认} \rightarrow \text{John}, \text{wanted} \rightarrow \text{John}
\end{align*}
\]

[Lewis et al, 2015]
Training

Learn latent CCG that recovers SRL

He opened the door

ARG0

ARG1
Training

Learn latent CCG that recovers SRL

• Generate *consistent* CCG/SRL parses for training sentences
Training

Learn latent CCG that recovers SRL

• Mark subset as correct, based on semantic dependencies
Training

Learn latent CCG that recovers SRL

- Optimize marginal likelihood
SRL Results

[Lewis et al 2015]
Out-of-domain SRL Results

F1 scores for different approaches:
- Riedel
- Zhao
- Che
- Vickrey
- Pipeline
- Joint

The Joint approach has the highest F1 score, followed by Vickrey, Che, and Pipeline. Riedel and Zhao have the lowest scores.
Towards Broad Coverage Semantic Parsing

- Can we crowdsource semantics?
- Train with latent syntax?
- Build fast and accurate parsers?
- Actively select which data to label?
Global A* Parsing

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

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

**Result:**
Accurate models with formal guarantees

[Lee et al, 2016, EMNLP best paper]
Fruit flies like bananas

Klein and Manning, 2001
Parsing with Hypergraphs

Fruit flies like bananas

Input

Output
Fruit flies like bananas

Each hyperedge $e$ is weighted with a score $g(e)$
Parsing with Hypergraphs

Score of parse derivation:

\[ g(y) = \sum_{e \in y} g(e) \]
Fruit flies like bananas

\[
\begin{align*}
\text{NP/ NP} & \rightarrow \text{NP} \\
\text{NP} & \rightarrow \text{NP} \\
\text{(S\(\backslash\)NP)/NP} & \rightarrow \text{S\(\backslash\)NP} \\
\text{S\(\backslash\)NP} & \rightarrow \text{NP}
\end{align*}
\]
Predicted parse: \( y^* = \arg\max_{y \in Y} g(y) \)

- Exponential number of nodes

\[ \rightarrow \] 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

Fruit flies like bananas

Dynamic programs with locally factored models, e.g.

Recursive neural networks break dynamic programs!

Scores condition on local structures

Fruit flies like bananas

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:**

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

- Efficient
- Inexpressive

**Global model:**

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

- Intractable
- Expressive
This Work

Combined model:

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

Efficient

Expressive
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

Klein and Manning, 2003
A* Parsing

Fruit flies like bananas

Partial parse
Fruit flies like bananas

A* Parsing

Partial parse
A* Parsing

\[ f ??? ? \]

Fruit flies like bananas

Partial parse

Exploration priority
A* Parsing

Exploration priority

Inside score

Admissible A* heuristic

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

### Agenda Position

<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</td>
</tr>
<tr>
<td>2</td>
<td>3.1</td>
<td>like (NP/NP)</td>
</tr>
<tr>
<td>3</td>
<td>1.9</td>
<td>Fruit</td>
</tr>
<tr>
<td>4</td>
<td>-0.5</td>
<td>Fruit (NP/NP)</td>
</tr>
</tbody>
</table>
A* Parsing

**Agenda position** | **f(y)** | **y**  
---|---|---
2 | 3.1 | like \((S\backslash NP)/NP\)  
3 | 1.9 | Fruit \(NP\)  
4 | -0.5 | Fruit \(NP/NP\)
A* Parsing

<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\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/NP</td>
</tr>
<tr>
<td>4</td>
<td>-1.3</td>
<td>flies NP</td>
</tr>
</tbody>
</table>

Fruit flies like bananas

---

Fruit flies like bananas
A* Parsing

<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(\text{NP}))/(\text{NP})</td>
</tr>
<tr>
<td>2</td>
<td>1.9</td>
<td>Fruit (\text{NP})</td>
</tr>
<tr>
<td>3</td>
<td>-0.5</td>
<td>Fruit (\text{NP}/\text{NP})</td>
</tr>
<tr>
<td>4</td>
<td>-1.3</td>
<td>flies (\text{NP})</td>
</tr>
</tbody>
</table>
A* Parsing

Agenda position | f(y) | y
--- | --- | ---
2 | 1.9 | Fruit \( NP \)
3 | -0.5 | Fruit \( NP/NP \)
4 | -1.3 | flies \( NP \)
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 (S\NP)/NP</td>
</tr>
<tr>
<td></td>
<td></td>
<td>bananas 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/\NP</td>
</tr>
<tr>
<td>4</td>
<td>-1.3</td>
<td>flies \NP</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>2.1</td>
<td>like $(S\backslash NP)/NP$ bananas $NP$</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 $NP/NP$</td>
</tr>
<tr>
<td>4</td>
<td>-1.3</td>
<td>flies $NP$</td>
</tr>
</tbody>
</table>
A* Parsing

**Table:**

<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 $\overline{NP}$</td>
</tr>
<tr>
<td>2</td>
<td>-1.5</td>
<td>like $\overline{(S\backslash S)/NP}$</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):

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

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

\[ g_{local} \left( \frac{NP/NP}{NP/NP} \right) + g_{local} \left( \frac{NP}{NP} \right) + g_{local} \left( \frac{(S\backslash NP)/NP}{NP} \right) + g_{local} \left( \frac{NP}{NP} \right) \]
Locally Factored Model

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

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

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

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

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

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

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

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

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

Non-positive global model

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

\[ h(y) = h_{\text{local}}(y) + 0 \]
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: $g_{global}(y)$

**Word embeddings**

**Bidirectional LSTM**

**Tree-LSTM**

**Parse Scores**

Diagram:

- **S**
- **NP**
- **NP/NP**
- **NP\NP**
- **NP/\NP**
- **NP**

Words:

- Fruit
- flies
- like
- bananas

Equation:

$$g_{global}(y)$$
Non-positive Global Model

\[ g_{global} = \log(\sigma(w \cdot \text{NP})) \]
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
Learning with A*
Learning with A*
Learning with A*
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</td>
<td>✗</td>
</tr>
<tr>
<td>2</td>
<td>-0.5</td>
<td>Fruit</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

$$A : \begin{bmatrix} \vdots & \cdots & \vdots \end{bmatrix}$$
Violation-based Loss

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

\( A : [ \ldots \]
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>$\times$</td>
</tr>
<tr>
<td>2</td>
<td>-0.5</td>
<td>Fruit $\frac{NP}{NP}$</td>
<td>$\checkmark$</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!
CCG Parsing Results

<table>
<thead>
<tr>
<th></th>
<th>Test F1 (%)</th>
<th>Is global?</th>
<th>Is exact?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clark &amp; Curran (2007)</td>
<td>85.2</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>Xu et al. (2015)</td>
<td>87.0</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Lewis et al. (2016)</td>
<td>88.1</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Vaswani et al. (2016)</td>
<td>88.3</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Global A*</td>
<td>88.7</td>
<td>✓</td>
<td>✓</td>
</tr>
</tbody>
</table>
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>
Garden Paths

Incorrect partial parse (syntactically plausible in isolation):

Input sentence:
The favorite **U.S. small business is one** whose research and development can be milked for future Japanese use.

Heavily penalized by the global model
Towards Broad Coverage Semantic Parsing

• Can we crowdsource semantics?
• Train with latent syntax?
• Build fast and accurate parsers?
• Actively select which data to label?
Our key hypothesis:
Anyone who **understands the meaning of a sentence** should be able to correct **parser mistakes**.

Pat ate the cake on the table that I **baked** last night.

Parser: I **baked** table

Human understanding: I **baked** cake

Can we use human judgements to improve parse?

[He et al, 2016]
Pat ate the cake on the table that I baked last night.

Q: What did someone bake?
   1. table    2. cake
Q: “What did someone bake?”
1) table 2) cake

Candidate dependencies from the n-best list:
baked → table
baked → cake

Re-parsed CCG Dependency Tree

C_pos (bake → cake)
C_neg (bake → table)

Not re-training the model
Generate Q/A Pairs from CCG Dependencies

Predicted CCG category of \textit{baked}: \((S\backslash NP_1)/NP_2\)

Convert to template: \begin{tabular}{c}
NP_1 \text{ bake} \text{ NP}_2 \\
\end{tabular}

Filling-in the Slots:

\begin{tabular}{c}
\textit{what} \text{ bake} \text{ sth.} \\
\end{tabular}

\begin{tabular}{c}
\textit{What} \text{ baked} \text{ something?} \\
\begin{tabular}{c}
\text{— I} \\
\end{tabular} \\
\end{tabular}

\begin{tabular}{c}
\textit{What} \text{ baked} \text{ something?} \\
\begin{tabular}{c}
\text{— I} \\
\end{tabular} \\
\end{tabular}

\begin{tabular}{c}
\textit{sth.} \text{ bake} \text{ what} \\
\end{tabular}

\begin{tabular}{c}
\textit{What} \text{ did} \text{ someone} \text{ bake?} \\
\begin{tabular}{c}
\text{— the table} \\
\end{tabular} \\
\end{tabular}

\begin{tabular}{c}
\textit{What} \text{ did} \text{ someone} \text{ bake?} \\
\begin{tabular}{c}
\text{— the cake} \\
\end{tabular} \\
\end{tabular}

Infer \textit{someone/something} and the \textit{answer spans} based on the n-best parses

Used “\textit{what}” for all questions
### Group Q/A Pairs into Queries

<table>
<thead>
<tr>
<th>Questions</th>
<th>Answers</th>
<th>Scores</th>
<th>Question Confidence</th>
<th>Answer Uncertainty (Entropy)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>What</strong> baked something?</td>
<td>I</td>
<td>1.0</td>
<td>1.0</td>
<td>0.0</td>
</tr>
<tr>
<td><strong>What</strong> did someone bake?</td>
<td>the table</td>
<td>0.7</td>
<td>1.0</td>
<td>0.88</td>
</tr>
<tr>
<td></td>
<td>the cake</td>
<td>0.3</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>What</strong> was baked something</td>
<td>the table</td>
<td>0.1</td>
<td>0.1</td>
<td>0.0</td>
</tr>
<tr>
<td>something?</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- The table shows questions, answers, scores, question confidence, and answer uncertainty (entropy).
- The question **What baked something?** has a high confidence and no uncertainty.
- The question **What did someone bake?** has a moderate confidence and a moderate uncertainty.
- The question **What was baked something something?** has a low confidence and no uncertainty.
- A non-sensical question mark is indicated.
- There is no uncertainty marked.
Our Annotation Task

Sentence:
Pat ate the cake on the table that I baked last night.

Question:
What did someone bake?

Check one or more
☐ the cake
☐ the table
☐ None of the above.

Comment:

* Crowdsourcing platform: https://www.crowdflower.com/.
Data Collection with Crowdsourcing

- All developments are done on CCG-Dev only.
- Less than 2 queries per sentence, for about 60% of the sentences.
- **Cost:** 46 cents per query.
- **Speed:** 200 queries per hour.
Inter-Annotator Agreement

- Agreement is computed only for matching the exact set of answers. i.e. (A, B) and (B) are considered disagreement.
- Unanimous agreement for over 40% of the queries.
- Over 90% absolute majority.
Putting our hypothesis to the test: How well does annotators’ human understanding align with the gold syntax?

- **Successes**: Long-range attachment decisions
- **Challenges**: Syntax-semantics mismatch
- **Use heuristics to fix the mismatch problems at re-parsing time.**
Temple also said Sea Containers’ plan raises numerous legal, regulatory, financial and fairness issues, but didn’t elaborate.

What didn’t elaborate something?

4  Temple

1  Sea Containers’ plan

0  None of the above.
Success - Coordination

To *avoid* these costs, and a possible default, immediate action is imperative.

What would something *avoid*?

4  **these costs**

3  **a possible default**

0  None of the above.
Kalipharma is a New Jersey-based pharmaceuticals concern that *sells* products under the Purepac label.

What *sells* something?

5 Kalipharma

None of the above.

- Syntax-semantics mismatch
- Also happens with pronouns and appositives.
- Some cases are heuristically fixed during reparsing.
Timex had requested duty-free treatment for many types of watches, covered by 58 different U.S. tariff classifications.

What would be covered?

0 Timex
0 duty-free treatment
0 None of the above.

2 many types of watches
3 watches

• Annotators tend to struggle with headedness.
• We add “disjunctive constraint”, forcing the re-parser to produce either of the two dependencies.
Re-Parsing with Crowdsourced Constraints

Q1: What did someone bake?
- votes(cake) = 4
- votes(table) = 1
- votes(None of the above) = 0

\[ y^{\text{new}} = \arg \max_y \text{base-parser-score}(y) \]
\[ -T^+ \times 1(baked \rightarrow \text{cake} \in y) \]
\[ -T^- \times 1(baked \rightarrow \text{table} \in y) \]

- Penalizes parses that disagree with crowdsourced judgments.
- Constraints are decomposed by dependencies.
- Thresholds and penalties are tuned on CCG-Dev.
Re-parsing Results (Labeled F1)

- Modest improvement due to syntax-semantics mismatch.
- Larger improvement on out-of-domain data.

Active, Ser133-phosphorylated CREB effects transcription of CRE-dependent genes via interaction with the 265-kDa …
Re-parsing Results

- Modified parse trees for about 10% of the sentences after incorporating human judgments.
- Larger gain on changed sentences.
- Changed sentences are “more difficult” on average.
Towards Broad Coverage Semantic Parsing

- Can we crowdsource semantics?
  - Yes, but need more than verbs….

- Train with latent syntax?
  - Yes, but must extend to QA supervision…

- Build fast and accurate parsers?
  - Yes, but need to extend to latent-variable case…

- Actively select which data to label?
  - Yes, but need to scale up…
Questions