Data Recombination for Neural Semantic Parsing

Robin Jia and Percy Liang
Stanford University
Semantic Parsing

What states border Texas?

And(State, NextTo(StateId(Texas)))

[New Mexico, Oklahoma, Arkansas, Louisiana]

Outline

- Neural Semantic Parser
- Data Recombination
- Discussion
Outline

• Neural Semantic Parser
• Data Recombination
• Discussion
Semantic Parsing

- Traditional semantic parsers are engineering-heavy
- Can we get good performance with a domain-general model?
Semantic Parsing

- Geoquery
  - Input: *what is the population of iowa?*
  - Output: 
    \[
    \texttt{\_answer ( A , ( \_population ( B , A ) , \\
    \_const ( B , \_stateid ( iowa ) ) ) )}
    \]

- ATIS
  - Input: *list all flights from chicago to milwaukee*
  - Output: 
    \[
    \texttt{( \_lambda \$0 e ( \_and ( \_flight \$0 ) ( \_from \\
    \$0 \text{chicago:}\_ci ) ( \_to \$0 \text{milwaukee:}\_ci ) ) )}
    \]

- Overnight
  - Input: *what restaurants have takeout*
  - Output: 
    \[
    \texttt{( call list ( call filter ( call getProperty \\
    ( call singleton en.restaurant ) ( string ! type ) ) \\
    ( string takeout ) ) )}
    \]

Sequence-to-sequence Models

- Machine Translation
  - Input: the blue house
  - Output: la maison bleue
- Syntactic Parsing
  - Input: the dog barked
  - Output: (S (NP DT NN )_{NP} (VP VBD)_{VP } )_{S}
- And many more...

Semantic Parsing

• Treat semantic parsing as a sequence-to-sequence task
• Use domain-general attention-based neural model

Sutskever et al. (2014), Bahdanau et al. (2014), Luong et al. (2015).
Neural Semantic Parser

\[ h_0^F \rightarrow h_1^F \rightarrow h_2^F \rightarrow h_3^F \rightarrow h_4^F \]

\[ h_1^B \leftarrow h_2^B \leftarrow h_3^B \leftarrow h_4^B \leftarrow h_5^B \]

what states border texas
Neural Semantic Parser

$h_1^F$, $h_2^F$, $h_3^F$, $h_4^F$

what, states, border, texas

$h_1^B$, $h_2^B$, $h_3^B$, $h_4^B$
Neural Semantic Parser

what  states  border  texas
Neural Semantic Parser

\[ h_0^F \rightarrow h_1^F \rightarrow h_2^F \rightarrow h_3^F \rightarrow h_4^F \]

\[ h_1^B \leftarrow h_2^B \leftarrow h_3^B \leftarrow h_4^B \leftarrow h_5^B \]

What states border texas

12
Neural Semantic Parser

what states border texas

\[ b_1 \rightarrow b_2 \rightarrow b_3 \rightarrow b_4 \rightarrow s_1 \]
Neural Semantic Parser

what     states  border  texas

\[ b_1 \quad b_2 \quad b_3 \quad b_4 \]

\[ s_1 \quad c_1 \]
Neural Semantic Parser

what states border texas

\( b_1 \quad b_2 \quad b_3 \quad b_4 \)

\( (\quad And \quad City \quad State \quad iowa \quad utah \quad \ldots \quad y_1 = \quad )\)
Neural Semantic Parser

what states border texas

\( b_1 \quad b_2 \quad b_3 \quad b_4 \)

\( (\quad ) \quad \text{And} \quad \text{City} \quad \text{State} \quad \text{iowa} \quad \text{utah} \quad \ldots \)

\( y_1 = \)
Neural Semantic Parser

what states border texas

\[ b_1 \quad b_2 \quad b_3 \quad b_4 \]

\[ s_1 \quad c_1 \]

\[ y_1 = \text{And} \]

\[ s_2 \]
Neural Semantic Parser

what states border texas

\[ y = \text{And} \]
city('alabama','al','birmingham',284413).
city('alabama','al','mobile',200452).
city('alabama','al','montgomery',177857).
city('alabama','al','huntsville',142513).
city('alabama','al','tuscaloosa',75143).
city('alaska','ak','anchorage',174431).
city('arizona','az','phoenix',789704).
city('arizona','az','tucson',330537).
city('arizona','az','mesa',152453).
city('arizona','az','tempe',106919).
city('arizona','az','glendale',96988).
city('arizona','az','scottsdale',88622).
city('arkansas','ar','little rock',158915).
city('arkansas','ar','fort smith',71384).
city('arkansas','ar','north little rock',64388).
city('california','ca','los angeles',2966850).
city('california','ca','san diego',875538).
city('california','ca','san francisco',678974).
city('california','ca','san jose',629442).
city('california','ca','long beach',361334).
city('california','ca','oakland',339337).
city('california','ca','sacramento',275741).
city('california','ca','anaheim',219311).
city('california','ca','fresno',218202).
city('california','ca','santa ana',203713).
city('california','ca','riverside',170876).
city('california','ca','huntington beach',170505).
city('california','ca','stockton',149779).
city('california','ca','glendale',139060).
city('california','ca','fremont',131945).
...
Rare Entities

• Rare entities pose a problem!
  • If we see an entity name we didn’t see during training, how can we even generate the right logical form?
• Solution: **Attention-based copying**

See also Gu et al. (2016), Gulcehre et al. (2016)
Neural Semantic Parser

\[ y = \text{And (State , NextTo (StateId (}} \]

\[ y_t = \text{(And City State iowa utah ... \text{what states border texas}} \]

\[ b_1 \quad b_2 \quad b_3 \quad b_4 \]
Neural Semantic Parser

\[ y = \text{And (State , NextTo (StateId (} \right) \text{And City State iowa utah ...} \right) \text{texas} \]
Implementation Details

• Training: maximize loglikelihood of correct logical form with SGD

• Test time
  • Decode with beam search
  • Add missing parentheses
  • Prune logical forms that result in execution error
## Results (Take 1)

<table>
<thead>
<tr>
<th>System</th>
<th>Geoquery</th>
<th>ATIS</th>
<th>Overnight</th>
</tr>
</thead>
<tbody>
<tr>
<td>Zettlemoyer and Collins (2007)</td>
<td></td>
<td>84.6</td>
<td></td>
</tr>
<tr>
<td>Kwiatkowski et al. (2010)</td>
<td>88.9</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Liang et al. (2011)*</td>
<td></td>
<td>91.1</td>
<td></td>
</tr>
<tr>
<td>Kwiatkowski et al. (2013)</td>
<td></td>
<td>89.0</td>
<td></td>
</tr>
<tr>
<td>Zhao and Huang (2015)</td>
<td>88.9</td>
<td>84.2</td>
<td></td>
</tr>
<tr>
<td>Wang et al. (2015)</td>
<td></td>
<td></td>
<td>58.8</td>
</tr>
<tr>
<td>Our RNN Model</td>
<td>85.0</td>
<td>76.3</td>
<td>75.8</td>
</tr>
</tbody>
</table>

How can we do better?

* Not directly comparable; used a seed lexicon for predicates.
Outline

- Neural Semantic Parser
- Data Recombination
- Discussion
Structural Regularities

• Given
  • what are the major cities in iowa?
    And(City, Major, LocatedIn(StateId(iowa)))

• We know how to parse
  • what are the major cities in texas?
    And(City, Major, LocatedIn(StateId(texas)))
Structural Regularities

• How do we build a neural model that respects compositional structural regularities?
Recombinant Examples

Dataset:

what are the major cities in iowa?
And(City, Major, LocatedIn(StateId( iowa )))

what are states that border texas?
And(State, NextTo(StateId( texas )))
what are the major cities in iowa?

And(City, Major, LocatedIn(StateId(iowa)))

Step 1: Apply high-precision alignment rules
what are the major cities in iowa?

And(City, Major, LocatedIn(StateId(iowa)))

Step 1: Apply high-precision alignment rules
Step 2: Infer types of aligned fragments
Recombinant Examples

**ROOT** → *(what are the major cities in STATE?*,
And(City, Major, LocatedIn(STATE)))*

**STATE** → *(iowa, StateId(iowa))*

Step 1: Apply high-precision alignment rules
Step 2: Infer types of aligned fragments
Step 3: Generate grammar rules by abstracting aligned fragments
Recombinant Examples

\[
\text{ROOT} \rightarrow (\text{what are states that border State?}, \\
\quad \text{And(State, NextTo(State)))}
\]
\[
\text{STATE} \rightarrow (\text{texas, StateId(texas)})
\]

Step 1: Apply high-precision alignment rules
Step 2: Infer types of aligned fragments
Step 3: Generate grammar rules by abstracting
aligned fragments
(Repeat for other examples)
Recombinant Examples

*what are the major cities in STATE?*

And(City, Major, LocatedIn(STATE))

Step 1: Apply high-precision alignment rules
Step 2: Infer types of aligned fragments
Step 3: Generate grammar rules by abstracting aligned fragments
Step 4: Combine rules from different examples to form new “recombinant” examples
Recombinant Examples

what are the major cities in texas?

And(City, Major, LocatedIn(StateId(texas)))

Step 1: Apply high-precision alignment rules
Step 2: Infer types of aligned fragments
Step 3: Generate grammar rules by abstracting aligned fragments
Step 4: Combine rules from different examples to form new “recombinant” examples
Data Recombination

Step 1: **Generate recombinant examples** using training data and prior knowledge about domain
Data Recombination

Initial data

Prior Knowledge

Recombinant Examples

Model

Step 2: Train on these examples to increase model’s awareness of task structure
Training the Model

• At each epoch, sample new recombinant examples from grammar
  • Choose production rules uniformly at random
• Train on recombinant examples plus original training examples in 1:1 ratio
## Results (Take 2)

<table>
<thead>
<tr>
<th>System</th>
<th>Geoquery</th>
<th>ATIS</th>
<th>Overnight</th>
</tr>
</thead>
<tbody>
<tr>
<td>Zettlemoyer and Collins (2007)</td>
<td>84.6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kwiatkowski et al. (2010)</td>
<td>88.9</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Liang et al. (2011)*</td>
<td>91.1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kwiatkowski et al. (2013)</td>
<td>89.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Zhao and Huang (2015)</td>
<td>88.9</td>
<td>84.2</td>
<td></td>
</tr>
<tr>
<td>Wang et al. (2015)</td>
<td>58.8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Our RNN Model</td>
<td>85.0</td>
<td>76.3</td>
<td>75.8</td>
</tr>
<tr>
<td>+ Abstracting Entities</td>
<td>85.4</td>
<td>79.9</td>
<td>75.3</td>
</tr>
</tbody>
</table>

* Not directly comparable; used a seed lexicon for predicates.
Structural Regularities, Part 2

• Given
  • what are the major cities in iowa?
    And(City, Major, LocatedIn(StateId(iowa)))
  • what are states that border texas?
    And(State, NextTo(StateId(texas)))

• We know how to parse
  • what are the major cities in states that border texas?
    And(City, Major, LocatedIn(And(State, NextTo(StateId(texas)))))
Abstracting Whole Phrases

What are states that border Texas?

And(State, NextTo(StateId(texas)))

Step 1: Apply high-precision alignment rules
What are states that border Texas?

And(State, NextTo(StateId(texas)))

Step 1: Apply high-precision alignment rules
Step 2: Infer types of aligned fragments
Abstracting Whole Phrases

What are states that border texas?

And\((State, \text{NextTo}(\text{StateId(texas)}))\)

Step 1: Apply high-precision alignment rules
Step 2: Infer types of aligned fragments

New Grammar Rule
- \(\text{STATE} \rightarrow (\text{states that border texas, And(State, NextTo(StateId(texas)))})\)
## Results (Take 3)

<table>
<thead>
<tr>
<th>System</th>
<th>Geoquery</th>
<th>ATIS</th>
<th>Overnight</th>
</tr>
</thead>
<tbody>
<tr>
<td>Zettlemoyer and Collins (2007)</td>
<td></td>
<td>84.6</td>
<td></td>
</tr>
<tr>
<td>Kwiatkowski et al. (2010)</td>
<td>88.9</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Liang et al. (2011)*</td>
<td></td>
<td>91.1</td>
<td></td>
</tr>
<tr>
<td>Kwiatkowski et al. (2013)</td>
<td></td>
<td>89.0</td>
<td></td>
</tr>
<tr>
<td>Zhao and Huang (2015)</td>
<td>88.9</td>
<td>84.2</td>
<td></td>
</tr>
<tr>
<td>Wang et al. (2015)</td>
<td></td>
<td>58.8</td>
<td></td>
</tr>
<tr>
<td>Our RNN Model</td>
<td>85.0</td>
<td>76.3</td>
<td>75.8</td>
</tr>
<tr>
<td>+ AbsEntities</td>
<td>85.4</td>
<td>79.9</td>
<td>75.3</td>
</tr>
<tr>
<td>+ AbsWholePhrases</td>
<td>87.5</td>
<td>---</td>
<td>75.9</td>
</tr>
</tbody>
</table>

* Not directly comparable; used a seed lexicon for predicates.
Composition of Strategies

• Grammar induction strategies are functions on grammars
  • Process each rule in grammar independently to generate new grammar rules
• Initial grammar
  • \texttt{ROOT} \rightarrow (\texttt{what are the major cities in iowa ?},
    \texttt{And(City, Major, LocatedIn(StateId(iowa))})
  • \texttt{ROOT} \rightarrow (\texttt{what are states that border texas ?},
    \texttt{And(State, NextTo(StateId(texas))})}
Composition of Strategies

Initial Grammar

Abstract Whole Phrases

AbsWholePhrases Grammar
STATE →

(states that border texas ,

And(State, NextTo(StateId(texas))))
Composition of Strategies

Initial Grammar

Abstract Whole Phrases

AbsWholePhrases Grammar

Abstract Entities

Composed Grammar
### Results (Take 4)

<table>
<thead>
<tr>
<th>System</th>
<th>Geoquery</th>
<th>ATIS</th>
<th>Overnight</th>
</tr>
</thead>
<tbody>
<tr>
<td>Zettlemoyer and Collins (2007)</td>
<td></td>
<td>84.6</td>
<td></td>
</tr>
<tr>
<td>Kwiatkowski et al. (2010)</td>
<td>88.9</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Liang et al. (2011)*</td>
<td>91.1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kwiatkowski et al. (2013)</td>
<td>89.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Zhao and Huang (2015)</td>
<td>88.9</td>
<td>84.2</td>
<td></td>
</tr>
<tr>
<td>Wang et al. (2015)</td>
<td></td>
<td></td>
<td>58.8</td>
</tr>
<tr>
<td>Our RNN Model</td>
<td>85.0</td>
<td>76.3</td>
<td>75.8</td>
</tr>
<tr>
<td>+ AbsEntities</td>
<td>85.4</td>
<td>79.9</td>
<td>75.3</td>
</tr>
<tr>
<td>+ AbsWholePhrases</td>
<td>87.5</td>
<td>---</td>
<td>75.9</td>
</tr>
<tr>
<td>+ AbsWholePhrases, AbsEntities</td>
<td>88.9</td>
<td>---</td>
<td>75.3</td>
</tr>
</tbody>
</table>

* Not directly comparable; used a seed lexicon for predicates.
What else can we do?

• What happens if we artificially make longer examples?
Concatenation

• Given
  • what are the major cities in iowa?
    And(City, Major, LocatedIn(StateId(iowa)))
  • what are states that border texas?
    And(State, NextTo(StateId(texas)))

• Create new example
  • what are the major cities in iowa? </s> what are states that border texas?
    And(City, Major, LocatedIn(StateId(iowa)))
    </s> And(State, NextTo(StateId(texas)))
Concatenation

- Can cast as grammar induction
- Compose with previous grammar induction strategies
## Final Results

<table>
<thead>
<tr>
<th>System</th>
<th>Geoquery</th>
<th>ATIS</th>
<th>Overnight</th>
</tr>
</thead>
<tbody>
<tr>
<td>Zettlemoyer and Collins (2007)</td>
<td>84.6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kwiatkowski et al. (2010)</td>
<td>88.9</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Liang et al. (2011)*</td>
<td>91.1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kwiatkowski et al. (2013)</td>
<td>89.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Zhao and Huang (2015)</td>
<td>88.9</td>
<td>84.2</td>
<td></td>
</tr>
<tr>
<td>Wang et al. (2015)</td>
<td></td>
<td></td>
<td>58.8</td>
</tr>
<tr>
<td>Our RNN Model</td>
<td>85.0</td>
<td>76.3</td>
<td>75.8</td>
</tr>
<tr>
<td>+ AbsEntities</td>
<td>85.4</td>
<td>79.9</td>
<td>75.3</td>
</tr>
<tr>
<td>+ AbsWholePhrases</td>
<td>87.5</td>
<td>---</td>
<td>75.9</td>
</tr>
<tr>
<td>+ AbsWholePhrases, AbsEntities</td>
<td>88.9</td>
<td>---</td>
<td>75.3</td>
</tr>
<tr>
<td>+ AbsWholePhrases, AbsEntities, Concat</td>
<td><strong>89.3</strong></td>
<td>83.3</td>
<td><strong>77.5</strong></td>
</tr>
</tbody>
</table>

* Not directly comparable; used a seed lexicon for predicates.
Outline

- Neural Semantic Parser
- Data Recombination
- Discussion
Data Recombination

Step 1: **Generate recombinant examples** using training data and prior knowledge about domain
Data Recombination

• Builds a **generative model** from the data
  • Mixture of the empirical distribution and induced probabilistic grammar
  • Train on examples \((x, y)\) sampled from this model
Data Augmentation

• Apply local transformations to a single example to generate more examples

\[ x = \text{image of a puppy} \]
\[ y = \text{“puppy”} \]

Krizhevsky et al. (2012).
Data Augmentation

- Apply local transformations to a single example to generate more examples

\[ x = \] 

\[ y = \text{“puppy”} \]

Horizontal Reflection

Krizhevsky et al. (2012).
Data Augmentation

• Apply local transformations to a single example to generate more examples

\[ x = \text{Cropping} \]

\[ y = \text{“puppy”} \]

Krizhevsky et al. (2012).
Data Recombination

• Get a richer generative model by combining information from **multiple examples**

what are the major cities in iowa?
And(City, Major, LocatedIn(StateId( iowa )))

what are states that border texas?
And(State, NextTo(StateId( texas )))
Empirical Distribution
Data Augmentation
Data Recombination
Thank you!

Initial data

Prior Knowledge

Recombinant Examples

Model

Code, data, and experiments available on CodaLab

https://worksheets.codalab.org/worksheets/0x50757a37779b485f89012e4ba03b6f4f/
Composition of Strategies

\[
\text{STATE} \rightarrow \text{(states that border texas,} \text{And(State, NextTo(StateId(texas))))}
\]

New Grammar Rule

- \text{STATE} \rightarrow (\text{states that border STATEID,} \text{And(State, NextTo(STATEID))))
An Artificial Experiment

• Make up some simple artificial data
  • *friends of relatives of alice*
    Friend(Relative(alice))
  • *colleagues of brothers of bob*
    Colleague(Brother(bob))

• Generate recombinant examples to train on
  • **Same length**: only swap entities
    • *friends of relatives of bob*
      Friend(Relative(bob))
  • **Longer**: nest whole phrases
    • *friends of relatives of colleagues of brothers of alice*
      Friend(Relative(Colleague(Brother(alice)))))
An Artificial Experiment

- Test only on “short” examples like those in the original training data
  - siblings of roommates of eve
    Sibling(Roommate(eve))
An Artificial Experiment

Test Accuracy

- No Data Recombination
- Data Recombination, Same-Length
- Data Recombination, Longer