Iterative Search for Weakly Supervised Semantic Parsing

Pradeep Dasigi
Matt Gardner
Shikhar Murty
Luke Zettlemoyer
Ed Hovy

Allen Institute for Artificial Intelligence
Mila
University of Washington
Carnegie Mellon University
This talk in one slide

- Training semantic parsing with denotation-only supervision is challenging because of **spuriousness**: incorrect logical forms can yield correct denotations.

- Two solutions:
  - Iterative training: Online search with initialization $\leftrightarrow$ MML over offline search output
  - Coverage during online search

- State-of-the-art single model performances:
  - WikiTableQuestions with comparable supervision
  - NLVR semantic parsing with significantly less supervision
## Semantic Parsing for Question Answering

<table>
<thead>
<tr>
<th>Athlete</th>
<th>Nation</th>
<th>Olympics</th>
<th>Medals</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gillis Grafström</td>
<td>Sweden (SWE)</td>
<td>1920–1932</td>
<td>4</td>
</tr>
<tr>
<td>Kim Soo-Nyung</td>
<td>South Korea (KOR)</td>
<td>1988-2000</td>
<td>6</td>
</tr>
<tr>
<td>Evgeni Plushenko</td>
<td>Russia (RUS)</td>
<td>2002–2014</td>
<td>4</td>
</tr>
<tr>
<td>Kim Yu-na</td>
<td>South Korea (KOR)</td>
<td>2010–2014</td>
<td>2</td>
</tr>
<tr>
<td>Patrick Chan</td>
<td>Canada (CAN)</td>
<td>2014</td>
<td>2</td>
</tr>
</tbody>
</table>

**Question:** Which athlete was from South Korea after the year 2010?

**Answer:** Kim Yu-Na

**Reasoning:**
1) Get rows where Nation is South Korea
2) Filter rows where value in Olympics > 2010.
3) Get value from Athlete column

**Program:**
```
(select_string
  (filter_in
    (filter > all_rows olympics 2010)
    south_korea)
  athlete)
```
Weakly Supervised Semantic Parsing

\( x_i \): Which athlete was from South Korea after 2010?

\( y_i \): (select_string (filter \( \triangleright \) all_rows olympics > 2010) south_korea) athlete

\( z_i \): Kim Yu-Na

\( w_i \):

<table>
<thead>
<tr>
<th>Athlete</th>
<th>Nation</th>
<th>Olympics</th>
<th>Medals</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kim Yu-na</td>
<td>South Korea</td>
<td>2010–2014</td>
<td>2</td>
</tr>
<tr>
<td>Tenley Albright</td>
<td>United States</td>
<td>1952-1956</td>
<td>2</td>
</tr>
</tbody>
</table>

Train on \( D = \{ x_i, w_i, z_i \}_{i=1}^{N} \)

Test: Given \( x_{N+k}, w_{N+k} \) find \( y_{N+k} \) such that \( [y_{N+k}]^{w_{N+k}} = z_{N+k} \)
### Challenge: Spurious logical forms

<table>
<thead>
<tr>
<th>Athlete</th>
<th>Nation</th>
<th>Olympics</th>
<th>Medals</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gillis Grafström</td>
<td>Sweden (SWE)</td>
<td>1920–1932</td>
<td>4</td>
</tr>
<tr>
<td>Kim Soo-Nyung</td>
<td>South Korea (KOR)</td>
<td>1988–2000</td>
<td>6</td>
</tr>
<tr>
<td><strong>Evgeni Plushenko</strong></td>
<td>Russia (RUS)</td>
<td>2002–2014</td>
<td>4</td>
</tr>
<tr>
<td><strong>Kim Yu-na</strong></td>
<td>South Korea (KOR)</td>
<td>2010–2014</td>
<td>2</td>
</tr>
<tr>
<td>Patrick Chan</td>
<td>Canada (CAN)</td>
<td>2014</td>
<td>2</td>
</tr>
</tbody>
</table>

**Which athletes are from South Korea after 2010?** Kim Yu-Na

**Logical forms that lead to answer:**

- Athlete from South Korea after 2010
- Athlete from South Korea with 2 medals
- First athlete in the table with 2 medals
- Athlete in row 4
Challenge: Spurious logical forms

There is exactly one square touching the bottom of a box

True

Logical forms that lead to answer:

- (count_equals(square(touch_bottom all_objects)) 1)
- (count_equals(yellow(square all_objects)) 1)
- (object_exists(yellow(triangle all_objects)))
- (object_exists all_objects)

Count of squares touching bottom of boxes is 1

Count of yellow squares is 1

There exists a yellow triangle

There exists an object

Due to binary denotations, 50% of logical forms give correct answer!
Training Objectives

**Maximum Marginal Likelihood**

- Eg.: Liang et al. (2011), Berant et al. (2013), Krishnamurthy et al. (2017, 2018), and others

\[
\max_{\theta} \prod_{x_i, w_i, z_i \in D} \sum_{y_i \in Y} p(y_i | x_i; \theta)
\]

... but we need a good set of approximate logical forms

**Reward/Cost-based approaches**

- Eg.: Neelakantan et al. (2016), Liang et al. (2017, 2018), and others

\[
\min_{\theta} \sum_{I=1}^{N} \mathbb{E}_{p(y_i | x_i; \theta)} C(x_i, y_i, w_i, d_i)
\]

Proposal: Alternate between the two objectives while gradually increasing the search space!

... but random initialization can cause the search to get stuck in the exponential search space!
Spuriousness solution 1: Iterative search

\[ D = \{x_i, w_i, z_i\}_{i=1}^N \]

Max logical form depth = k

\[ D^0 = \{x_j, Y_j\}_{j=1}^M \]

\[ \forall y_j \in Y_j C(x_j, y_j, w_j, d_j) = 0 \]

Step 0: Get seed set of logical forms till depth k
Spuriousness solution 1: Iterative search

\[ D = \{x_i, w_i, z_i\}_{i=1}^{N} \]

**Limited depth exhaustive search**

Max logical form depth = k

\[ D^0 = \{x_j, Y_j\}_{j=1}^{M} \]

\[ \forall y_j \in Y_j C(x_j, y_j, w_j, d_j) = 0 \]

**Step 0:** Get seed set of logical forms till depth k

**Step 1:** Train model using MML on seed set

Maximum Marginal Likelihood
Spuriousness solution 1: Iterative search

\[ D = \{x_i, w_i, z_i\}_{i=1}^N \]

**Step 0:** Get seed set of logical forms till depth \( k \)

**Step 1:** Train model using MML on seed set

**Step 2:** Train using MBR on all data till a greater depth \( k + s \)
Spuriousness solution 1: Iterative search

\[ D = \{x_i, w_i, z_i\}_{i=1}^N \]

\[ D^1 = \{x_l, Y_l\}_{l=1}^P \]

\[ \forall y_l \in Y_l \, C(x_l, y_l, w_l, d_l) = 0 \]

Max logical form depth = \( k + s \)

Minimum Bayes Risk training till depth \( k + s \)

**Step 0:** Get seed set of logical forms till depth \( k \)

**Step 1:** Train model using MML on seed set

**Step 2:** Train using MBR on all data till a greater depth \( k + s \)

**Step 3:** Replace offline search with trained MBR and update seed set
Spuriousness solution 1: Iterative search

\[ D = \{x_i, w_i, z_i\}_{i=1}^{N} \]

\[ D^1 = \{x_l, Y_l\}_{l=1}^{P} \]

\[ \forall y_l \in Y_l C(x_l, y_l, w_l, d_l) = 0 \]

Step 0: Get seed set of logical forms till depth \( k \)

Step 1: Train model using MML on seed set

Step 2: Train using MBR on all data till a greater depth \( k + s \)

Step 3: Replace offline search with trained MBR and update seed set

\[ k : k + s; \text{ Go to Step 1} \]

Iterate till dev. accuracy stops increasing

Maximum Marginal Likelihood

LSTM \[\rightarrow\] LSTM \[\rightarrow\] LSTM \[\rightarrow\] LSTM
Spuriousness Solution 2: Coverage guidance

There is exactly one square touching the bottom of a box.

(count_equals (square (touch_bottom all_objects)) 1)

- **Insight:** There is a significant amount of trivial overlap
- **Solution:** Use overlap as a measure guide search
There is exactly one square touching the bottom.

Target symbols triggered by rules:
- count_equals
- square
- touch_bottom

Coverage cost is the number of triggered symbols that do not appear in the logical form:

Lexicon:
- there is a box → box_exists
- there is a [other] → object_exists
- box … blue → color_blue
- box … black → color_black
- box … yellow → color_yellow
- box … square → shape_square
- box … circle → shape_circle
- box … triangle → shape_triangle
- not → negate_filter
- contains → object_in_box
- touch … top → touch_top
- touch … bottom → touch_bottom
- touch … corner → touch_corner
- touch … right → touch_right
- touch … left → touch_left
- touch … wall → touch_edge

Example: There is exactly one square touching the bottom of a box.

Triggered target symbols: \{count\_equals, square, 1, touch\_bottom\}

Coverage costs of candidate logical forms:

<table>
<thead>
<tr>
<th>Logical form</th>
<th>Coverage</th>
</tr>
</thead>
<tbody>
<tr>
<td>(count_equals (square (touch_bottom all_objects) 1))</td>
<td>0</td>
</tr>
<tr>
<td>(count_equals (square all_objects) 1)</td>
<td>1</td>
</tr>
<tr>
<td>(object_exists all_objects)</td>
<td>4</td>
</tr>
</tbody>
</table>
Training with Coverage Guidance

- Augment the reward-based objective:

$$\min_{\theta} \sum_{i=1}^{N} \mathbb{E}_{p(y_i|x_i;\theta)} C(x_i, y_i, w_i, d_i)$$

Now, $C$ is defined a linear combination of **coverage** and **denotation** costs

$$C(x_i, y_i, w_i, d_i) = \lambda S(y_i, x_i) + (1 - \lambda) T(y_i, w_i, d_i)$$
Results of training with iterative search on NLVR*

* using structured representations
Results of training with iterative search on WikiTableQuestions

![Bar chart showing development accuracy for different iterations with MBR Acc and MML Acc.]

- Iteration 0: MBR Acc = 40, MML Acc = 42.5
- Iteration 1: MBR Acc = 42.5, MML Acc = 42.5
- Iteration 2: MBR Acc = 43.1, MML Acc = 42.7
- Iteration 3: MBR Acc = 42.8, MML Acc = 42.5
Results of using coverage guided training on NLVR

* using structured representations

- Model does not learn without coverage!
  - Majority baseline: 56.2
  - MBR w/o coverage: 56.4
  - MBR w/ coverage: 73.9

- Coverage helps even with strong initialization
  - MBR + MML Init: 77.7
  - MBR + MML Init + coverage: 80.7

when trained from scratch

when model initialized from an MML model trained on a seed set of offline searched paths
Comparison with previous approaches on NLVR*

- MaxEnt, BiAttPonter are not semantic parsers
- Abs. supervision + Rerank uses manually labeled abstractions of utterance - logical form pairs to get training data for a supervised system, and reranking
- Our work outperforms Goldman et al., 2018 with fewer resources

* using structured representations
Comparison with previous approaches on WikiTableQuestions

<table>
<thead>
<tr>
<th>Category</th>
<th>Model</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Non-neural models</td>
<td>Pasupat and Liang (2015)</td>
<td>37.1</td>
</tr>
<tr>
<td></td>
<td>Zhang et al. (2017)</td>
<td>43.7</td>
</tr>
<tr>
<td>Reinforcement Learning</td>
<td>Neelakantan et al. (2017)</td>
<td>34.2</td>
</tr>
<tr>
<td></td>
<td>Liang et al. (2018) (avg)</td>
<td>43.1</td>
</tr>
<tr>
<td></td>
<td>Liang et al. (2018) (best)</td>
<td>43.8</td>
</tr>
<tr>
<td>Non-RL Neural Models</td>
<td>Haug et al. (2018)</td>
<td>34.8</td>
</tr>
<tr>
<td></td>
<td>This work (avg)</td>
<td>43.9</td>
</tr>
<tr>
<td></td>
<td>This work (best)</td>
<td>44.3</td>
</tr>
</tbody>
</table>
Summary

- Spuriousness is a challenge in training semantic parsers with weak supervision
- Two solutions:
  - Iterative training: Online search with initialization ⇆ MML over offline search output
  - Coverage during online search
- SOTA single model performances:
  - WikiTableQuestions: 44.3%
  - NLVR semantic parsing: 82.9%

Thank you!
Questions?