Iterative Search for Weakly Supervised Semantic Parsing

Pradeep Dasigi  Matt Gardner  Shikhar Murty  Luke Zettlemoyer  Ed Hovy
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

| Athlete         | Nation           | Olympics      | Medals |
|-----------------|------------------|---------------|--------|
| Gillis Grafström| Sweden (SWE)     | 1920–1932     | 4      |
| Kim Soo-Nyung   | South Korea (KOR)| 1988-2000     | 6      |
| Evgeni Plushenko| Russia (RUS)     | 2002–2014     | 4      |
| Kim Yu-na       | South Korea (KOR)| 2010–2014     | 2      |
| Patrick Chan    | Canada (CAN)     | 2014          | 2      |

**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)
```

WikiTableQuestions, Pasupat and Liang (2015)
Weakly Supervised Semantic Parsing

**$x_i$:** Which athlete was from South Korea after 2010?

**$y_i$:** $(\text{select_string}(\text{filter}_{\in}(\text{filter}_{> \text{all_rows olympics}}2010)\text{south_korea})\text{athlete})$

**$z_i$:** Kim Yu-Na

**$w_i$:**

| Athlete       | Nation      | Olympics     | Medals |
|---------------|-------------|--------------|--------|
| Kim Yu-na     | South Korea | 2010–2014    | 2      |
| Tenley Albright | United States | 1952-1956 | 2      |

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

| Athlete            | Nation            | Olympics       | Medals |
|--------------------|-------------------|----------------|--------|
| Gillis Grafström   | Sweden (SWE)      | 1920–1932      | 4      |
| Kim Soo-Nyung      | South Korea (KOR) | 1988–2000      | 6      |
| Evgeni Plushenko   | Russia (RUS)      | 2002–2014      | 4      |
| **Kim Yu-na**      | South Korea (KOR) | 2010–2014      | 2      |
| Patrick Chan       | Canada (CAN)      | 2014           | 2      |

**Logical forms that lead to answer:**

- Athlete from South Korea after 2010? Kim Yu-Na
- 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

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

Count of yellow squares is 1
There exists a yellow triangle
There exists an object

Cornell Natural Language Visual Reasoning, Suhr et al., 2017
Training Objectives

Maximum Marginal Likelihood
Maximize the marginal likelihood of an approximate set of logical forms

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

... 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

Minimum Bayes Risk training
Minimize the expected value of a cost
... but random initialization can cause the search to get stuck in the exponential search space

Proposal: Alternate between the two objectives while gradually increasing the search space!
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 \( k \)

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

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

\[ D^0 = \{x_j, Y_j\}_{j=1}^{M} \]
\[ \forall y_j \in Y_j C(x_j, y_j, w_j, d_j) = 0 \]
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 \( k \times s \)

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

- **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; Go to Step 1

Iterate till dev. accuracy stops increasing
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
Spuriousness Solution 2: Coverage guidance

There is exactly one square touching the bottom.

Target symbols triggered by rules:
count_equals
1
square
touch_bottom

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

Example:

Sentence: **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:

| Logical form                                                                 | Coverage |
|-----------------------------------------------------------------------------|----------|
| (count\_equals (square (touch\_bottom all\_objects)) 1)                      | 0        |
| (count\_equals (square all\_objects) 1)                                      | 1        |
| (object\_exists all\_objects)                                               | 4        |
Training with Coverage Guidance

- Augment the reward-based objective:

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

now \( \mathcal{C} \) is defined a linear combination of \textit{coverage} and \textit{denotation} costs

\[
\mathcal{C}(x_i, y_i, w_i, d_i) = \lambda \mathcal{S}(y_i, x_i) + (1 - \lambda) \mathcal{T}(y_i, w_i, d_i)
\]
Results of training with iterative search on NLVR* using structured representations

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

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

- Iteration 0: MBR Acc 40, MML Acc 40
- 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*

Model does not learn without coverage!

Coverage helps even with strong initialization

when trained from scratch

when model initialized from an MML model trained on a seed set of offline searched paths

* using structured representations
Comparison with previous approaches on NLVR* using structured representations

- 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
Comparison with previous approaches on WikiTableQuestions

Non-neural models

- Pasupat and Liang (2015): 37.1
- Zhang et al. (2017): 43.7

Reinforcement Learning models

- Neelakantan et al. (2017): 34.2
- Liang et al. (2018) (avg): 43.1
- Liang et al. (2018) (best): 43.8

Non-RL Neural Models

- Haug et al. (2018): 34.8
- This work (avg): 43.9
- This work (best): 44.3
Summary

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

Thank you!
Questions?