Weakly Supervised Semantic Parsing with Abstract Examples

Omer Goldman, Veronica Latcinnik, Udi Naveh, Amir Globerson, Jonathan Berant
Tel Aviv University
ACL, July 2018
**Semantic Parsing**

What is the capital of the largest US state?

\[ \text{CapitalOf}.\text{argmax}(\text{Type}.\text{State} \cap \text{LocatedIn}.\text{US}, \text{Population}) \]

Sacramento

KB:
Training with Full Supervision

- Training examples:

\[
x: \quad y: \text{CapitalOf.\arg\max_{\text{Type}} \text{State} \cap \text{LocatedIn.US, Population}}
\]

Introduction ➤ Semantic parser ➤ Abstract examples ➤ Results ➤ Conclusions
Training with Weak Supervision

Training examples:

\[ x: \]
\[ y: Sacramento \]
Problems with Weak Supervision

- Exponential search space

![Diagram showing exponential search space with examples: 3+3*15, 1, 5+30, 2-2, 4+60/3, with all resulting in incorrect decoding.](image)
Problems with Weak Supervision

- Spurious programs (Pasupat and Liang, 2016; Guu et al., 2017)

![Diagram showing decoding process with examples]

Correct program: $2 \times 2$
**CNLVR** (Suhr et al., 2017)

\[ I : \]

\[ k : [[[x_{-}loc: \ldots, color: 'Black', type: 'square', x_{-}loc: \ldots, size: 20}, \ldots]]] \]

\[ x : 'There is a small yellow item not touching any wall' \]

\[ y : True \]
There is a blue square

\[
\exists (\text{filter}(\text{ALL_ITEMS}, \lambda x. \text{IsBlue}(x) \land \text{IsSquare}(x)))
\]

Binary! True

50% spurious

KB:
Insight

There is exactly one black circle not touching the edge

- Equal(1, (filter(ALL_ITEMS, λx. IsBlack(x) ∩ IsCircle(x) ∩ ¬IsTouchingWall(x)))
- GreaterEqual(3, (filter(ALL_ITEMS, λx. IsBlue(x) ∩ IsTriangle(x) ∩ ¬IsTouchingWall(x)))
- GreaterEqual(1, (filter(ALL_ITEMS, λx. IsBlue(x) ∩ IsTriangle(x) ∩ ¬IsTouchingWall(x)))
- LessEqual(3, (filter(ALL_ITEMS, λx. IsYellow(x) ∩ IsRectangle(x) ∩ ¬IsTouchingWall(x)))
 Contributions

There is a yellow circle

exist(filter(ALL_ITEMS, \( \lambda x. \text{IsYellow}(x) \land \text{IsCircle}(x) \)))

There is a C–COLOR C–SHAPE

exist(filter(ALL_ITEMS, \( \lambda x. \text{IsC-COLOR}(x) \land \text{IsC-SHAPE}(x) \)))

Data augmentation helps search

Abstract cache tackles spuriousness

CNLVR improvement:

67.8 \rightarrow 82.5
Semantic Parsing
Logical Program

\[ x: \]

\[ z: \text{exist}(\text{filter}(\text{ALL\_BOXES}, \lambda x. \text{ge}(3, \text{count}(\text{filter}(x, \lambda y. \text{IsBlue}(y))))) ) \]
Model

Training maximizes log-likelihood of correct programs

+ discriminative re-ranker
Abstract Examples
Abstraction

There is a yellow circle
exist(filter(ALL_ITEMS, λ. IsYellow(x) ∧ IsCircle(x)))

There is a blue square
exist(filter(ALL_ITEMS, λ. IsBlue(x) ∧ IsRectangle(x)))

<table>
<thead>
<tr>
<th>Utterance</th>
<th>Program</th>
<th>Cluster</th>
<th>#</th>
</tr>
</thead>
<tbody>
<tr>
<td>&quot;yellow&quot;</td>
<td>IsYellow</td>
<td>C-Color</td>
<td>3</td>
</tr>
<tr>
<td>&quot;big&quot;</td>
<td>IsBig</td>
<td>C-Size</td>
<td>3</td>
</tr>
<tr>
<td>&quot;square&quot;</td>
<td>IsSquare</td>
<td>C-Shape</td>
<td>4</td>
</tr>
<tr>
<td>&quot;3&quot;</td>
<td>3</td>
<td>C-Num</td>
<td>2</td>
</tr>
<tr>
<td>&quot;exactly&quot;</td>
<td>EqualInt</td>
<td>C-QuantMod</td>
<td>5</td>
</tr>
<tr>
<td>&quot;top&quot;</td>
<td>Side.Top</td>
<td>C-Location</td>
<td>2</td>
</tr>
<tr>
<td>&quot;above&quot;</td>
<td>GetAbove</td>
<td>C-SpaceRel</td>
<td>6</td>
</tr>
<tr>
<td><strong>Total:</strong></td>
<td></td>
<td><strong>25</strong></td>
<td></td>
</tr>
</tbody>
</table>

There is a \(C – \text{COLOR} \& C – \text{SHAPE}\)
exist(filter(ALL_ITEMS, λ. IsC-COLOR(x) ∧ IsC-SHAPE(x)))
Analysis

3163 CNLVR sentences
  • There is............
  • One of the.......  
  • There are........
  • Exactly two.....
  • There is........
  • In two of.......  
  • There is........
  • There are......  
  • There is........
  • One square.....
  • There is........
  • One of the......
  • There are......  
  • There is........
  • Two towers.....
  • There are......  
  • There is........
  • One circle......
  • There is........
  • Last one........

~1300 abstract sentences
  • There is............
  • There are........
  • C-Num of.......... 
  • There is.........
  • One tower.........
  • There are........
  • C-Num C-Shape...
  • There is............
  • C-Num towers....
  • Another last........

~150 abstract sentences cover 50% of CNLVR.
Abstraction

- Data augmentation
- Abstract cache
There is a yellow circle

There is a blue rectangle
exist(filter(ALL_ITEMS, λx. IsBlue(x) ∧ IsSquare(x)))

There is a yellow triangle
exist(filter(ALL_ITEMS, λx. IsYellow(x) ∧ IsTriangle(x)))

There is a black circle
exist(filter(ALL_ITEMS, λx. IsBlack(x) ∧ IsCircle(x)))

There is a C – COLOR C – SHAPE
exist(filter(ALL_ITEMS, λx. IsC-COLOR(x) ∧ IsC-SHAPE(x)))
Training Procedure

~100 Abstract examples
(abs. sent., abs. prog.)

~6000 Instantiated examples
(sentence, program)

3163 CNLVR training examples
(sentence, answer)

Supervised model

Weakly-supervised model

Introduction ➞ Semantic parser ➞ Abstract examples ➞ Results ➞ Conclusions
Abstract Cache

“There is a yellow triangle”

abstraction

decoding

“There is a C-Color C-Shape”

retrieval

Beam:

\[
\begin{array}{cccccc}
t = 0 & t = 1 & t = 2 & t = 3 & t = 4 & t = 5 \\
\begin{aligned}
\text{Exist} & \quad \text{Filter} & \quad \text{ALL\_ITEMS} & \quad \lambda x.\text{And} & \quad \text{Count} & \quad \text{Filter} \\
\text{Exist} & \quad \text{Filter} & \quad \text{ALL\_ITEMS} & \quad \lambda x.\text{And} & \quad \text{IsYellow} & \quad \text{IsYellow} \\
\text{GreaterThan} & \quad 1 & \quad \text{Count} & \quad \text{Filter} & \quad \text{ALL\_ITEMS} & \quad \lambda x.\text{Equal} \\
\text{Equal} & \quad 2 & \quad \text{Count} & \quad \text{Filter} & \quad \text{ALL\_ITEMS} & \quad \lambda x.\text{Equal} \\
\text{Exist} & \quad \text{Filter} & \quad \text{ALL\_ITEMS} & \quad \lambda x.\text{And} & \quad \text{IsYellow} & \quad \text{IsTriangle} \\
\text{GreaterThan} & \quad 1 & \quad \text{Count} & \quad \text{Filter} & \quad \text{ALL\_ITEMS} & \quad \lambda x.\text{And} \\
\end{aligned}
\end{array}
\]

Exist(\text{Filter}(\text{ALL\_ITEMS}, \lambda x.\text{And}(\text{IsC\_COLOR}(x), \text{IsC\_SHAPE}(x)))) 95%

GreaterThan(1, \text{count}(\text{Filter}(\text{ALL\_ITEMS}, \lambda x.\text{And}(\text{IsC\_COLOR}(x), \text{IsC\_SHAPE}(x))))) 90%

Exist(\text{Filter}(\text{ALL\_ITEMS}, \lambda x.\text{IsCOLOR}(x))) 30%
Reward Tying

Introduction

Semantic parser

Abstract examples

Results

Conclusions

50% spurious

\[ I : \]

\[ k : \text{[[\{y-loc: \ldots, color: 'Black', type: 'square', x-loc: \ldots, size: 20}, \ldots\}]]} \]

\[ x : \text{There is a small yellow item not touching any wall} \]

\[ y : \text{True} \]
Reward Tying

\[ z : \text{Exist}(\text{Filter(ALL_ITEMS, } \lambda x. \text{And(And(IsYellow}(x), \text{IsSmall}(x)), \text{Not(IsTouchingWall}(x, \text{Side.Any})))))) \]
Results
Models

- Majority label (True)
- Max Entropy classifier on extracted features
- Supervised trained model (+Re-ranker)
- Weakly supervised trained model (+Re-ranker)

Baselines (Suhr et al., 2017)
Results - Public test set

<table>
<thead>
<tr>
<th>Method</th>
<th>Test-P Accuracy</th>
<th>Test-P Consistency</th>
</tr>
</thead>
<tbody>
<tr>
<td>Majority</td>
<td>56.2</td>
<td></td>
</tr>
<tr>
<td>MaxEnt</td>
<td>67.7</td>
<td></td>
</tr>
<tr>
<td>Sup.</td>
<td>66.9</td>
<td></td>
</tr>
<tr>
<td>Sup.+Rerank</td>
<td>76.6</td>
<td>51.8</td>
</tr>
<tr>
<td>W.Sup.</td>
<td>81.7</td>
<td>60.1</td>
</tr>
<tr>
<td>W.Sup.+Rerank</td>
<td>84</td>
<td>65</td>
</tr>
</tbody>
</table>

Test-P Accuracy and Test-P Consistency for different methods on the public test set.
Ablations

- No data augmentation
- Abstract weakly supervised parser
Ablations

- Abstraction
- Data augment.
- Beam cache
- W.Sup.+Rerank

- Dev Accuracy
- Dev Consistency

Data augmentation addition

Cache addition

Introduction  Semantic parser  Abstract examples  Results  Conclusions
Conclusions
Conclusions

Similar ideas in: Dong and Lapata (2018) and Zhang et al. (2017)

Automation would be useful
Thank you

https://github.com/udiNaveh/nlvr_tau_nlp_final_proj