Language to Logical Form with Neural Attention
August 8, 2016

Li Dong and Mirella Lapata
Transform **natural language** to **logical form**

Human friendly -> computer friendly

What is the highest mountain in Alaska?

\[ \text{argmax} \; 0 \; ( \text{and} \; (\text{mountain}:t \; 0) \; (\text{loc}:t \; 0 \; \text{alaska}:s) \; (\text{elevation}:i \; 0)) \]

* Example from GeoQuery
(natural language, logical form) pairs

(Miller et al., 1996; Zelle and Mooney, 1996; Tang and Mooney, 2000; Thomspn and Mooney, 2003; Kate et al., 2005; Ge and Mooney, 2005; Kate and Mooney, 2006; Wong and Mooney, 2006; Wong and Mooney, 2007; Zettlemoyer and Collins, 2005; 2007; Lu et al., 2008; Kwiatkowski et al., 2010; 2011; Andreas et al., 2013; Zhao and Huang, 2015;)

(natural language, answer) pairs

(Clarke et al., 2010; Artzi and Zettlemoyer, 2011; Chen and Mooney, 2011; Goldwasser and Roth, 2011; Artzi and Zettlemoyer, 2013; Liang et al., 2013; Krishnamurthy and Mitchell, 2012; Cai and Yates, 2013; Reddy et al., 2014;)

What is the highest mountain in Alaska ?

(argmax $0 (and (mountain:t $0) (loc:t $0 alaska:s)) (elevation:i $0))
Manually designed features
Predefined templates
Lexicon seeds

-> Domain- or representation-specific
Research Goal

- Reduce reliance on domain knowledge
- Use NNs to replace manually designed features
- Build a general-purpose parser: easy to adapt across domains and meaning representations
what microsoft jobs do not require a bscs?

Input Utterance

Sequence Encoder

Sequence/Tree Decoder

Logical Form

answer(J,(company(J,'microsoft'),job(J),not((req_degree(J,'bscs')))))

(Kalchbrenner and Blunsom, 2013; Cho et al., 2014; Sutskever et al., 2014; Karpathy and Fei-Fei, 2015; Vinyals et al., 2015;)

Language to Logical Form with Neural Attention 6 / 25
Sequence-to-Sequence (Seq2Seq) Model

Input: $q = x_1 \ldots x_{|q|}$

Encoder

Decoder

Output: $a = y_1 \ldots y_{|a|}$
Ignore the hierarchical structure of logical forms

Use `( )` to linearize logical form

\[
\lambda \, e \, \langle n \rangle \\
\text{and} \, \langle n \rangle \, \langle n \rangle \\
\left( > \, \langle n \rangle \, 1600:ti \right) \, \text{from} \, \langle n \rangle \, \langle n \rangle \\
\text{departure}_{-} time \, \langle n \rangle \\
\lambda \, e \, \left( \text{and} \, \left( > \, \left( \text{departure}_{-} time \, \langle n \rangle \right) \, 1600:ti \right) \, \text{from} \, \langle n \rangle \, \langle n \rangle \right)
\]
Drawback of Seq2Seq Model

Ignore the hierarchical structure of logical forms
More long-distance dependency during decoding

\[
\lambda \, e \, (\text{and} \, (\text{departure\_time} \, e) \, > \, 1600:ti) \, (\text{from} \, e \, dallas:ci)
\]
Define a “nonterminal” \(<n>\) token to indicate subtrees in decoder.
Seq2Tree Decoder

Language to Logical Form with Neural Attention
Attention Mechanism – Soft Alignment

Language to Logical Form with Neural Attention

(Bahdanau et al., 2015; Luong et al., 2015b; Xu et al., 2015)
**Training and Inference**

- **Training**
  
  \[
  \text{maximize} \sum_{(q,a) \in D} \log p(a|q)
  \]

- **Inference**
  
  \[
  \hat{a} = \arg\max_{a'} p(a'|q)
  \]
  
  Greedy/Beam search
Many questions contain entities or numbers

- Unavoidably rare
- Or do not appear in the training set at all

\[
\text{jobs with a salary of } <\text{unk}>.
\]

\[
\text{job}(\text{ANS}), \text{salary greater than } <\text{unk}>, \text{year}
\]

Replace rare words with \(<\text{unk}>\) (Luong et al., 2015; Jean et al., 2015)

Detrimental for semantic parsing
Argument Identification

- Pre-process entities and numbers to $type_i$
- At inference time, a post-processing step recovers maskers to their corresponding logical constants

$jobs$ with a salary of $num_0$

$job(ANS)$, salary greater than $(ANS, num_0, year)$
## Experiments

<table>
<thead>
<tr>
<th>Length</th>
<th>Jobs</th>
</tr>
</thead>
<tbody>
<tr>
<td>9.80</td>
<td>what microsoft jobs do not require a bscs?</td>
</tr>
<tr>
<td>22.90</td>
<td>answer(J,company(J,’microsoft’),job(J),not((req deg(J,’bscs’))))</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Length</th>
<th>Geo</th>
</tr>
</thead>
<tbody>
<tr>
<td>7.60</td>
<td>what is the population of the state with the largest area?</td>
</tr>
<tr>
<td>19.10</td>
<td>(argmax $0 (and (mountain:t $0) (loc:t $0 alaska:s)) (elevation:i $0))</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Length</th>
<th>ATIS</th>
</tr>
</thead>
<tbody>
<tr>
<td>11.10</td>
<td>dallas to san francisco leaving after 4 in the afternoon please</td>
</tr>
<tr>
<td>28.10</td>
<td>(lambda $0 e (and (&gt; (departure time $0) 1600:ti) (from $0 dallas:ci) (to $0 san francisco:ci)))</td>
</tr>
</tbody>
</table>
IFTTT (Quirk et al., 2015)

IF-This-Then-That

- turn on my lights when I arrive home
- text me if the door opens
- remind me to drink water if I’ve been at a bar for more than two hours

![Diagram showing IF-This-Then-That logic]

Archive your missed calls from Android to Google Drive
Experimental Results

Language to Logical Form with Neural Attention

### Jobs

<table>
<thead>
<tr>
<th>Method</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tang and Mooney, 2001</td>
<td>79.4</td>
</tr>
<tr>
<td>Popescu et al., 2003</td>
<td>88</td>
</tr>
<tr>
<td>Zettlemoyer and Collins, 2005</td>
<td>79.3</td>
</tr>
<tr>
<td>Liang et al., 2013</td>
<td>90.7</td>
</tr>
<tr>
<td>Zhao and Huang, 2015</td>
<td>85</td>
</tr>
<tr>
<td>vanilla Seq2Seq</td>
<td>70.7</td>
</tr>
<tr>
<td>w/argument</td>
<td>77.9</td>
</tr>
<tr>
<td>w/attention</td>
<td>87.1</td>
</tr>
<tr>
<td>Seq2Tree</td>
<td>90</td>
</tr>
</tbody>
</table>
Experimental Results

Geo

<table>
<thead>
<tr>
<th>Zettlemoyer and Collins, 2006</th>
<th>Kwiatkowski et al., 2010</th>
<th>Kwiatkowski et al., 2011</th>
<th>Liang et al., 2013</th>
<th>Zhao and Huang, 2015</th>
<th>vanilla Seq2Seq</th>
<th>w/argument</th>
<th>w/attention</th>
<th>Seq2Tree</th>
</tr>
</thead>
<tbody>
<tr>
<td>79.3</td>
<td>86.1</td>
<td>87.9</td>
<td>88.6</td>
<td>89</td>
<td>87.9</td>
<td>88.9</td>
<td>84.6</td>
<td>87.1</td>
</tr>
</tbody>
</table>

IFTTT

<table>
<thead>
<tr>
<th>retrieval</th>
<th>phrasal</th>
<th>sync</th>
<th>classifier</th>
<th>posclass</th>
<th>vanilla Seq2Seq</th>
<th>w/argument</th>
<th>w/attention</th>
<th>Seq2Tree</th>
</tr>
</thead>
<tbody>
<tr>
<td>56.2</td>
<td>45.5</td>
<td>42.8</td>
<td>65</td>
<td>66.5</td>
<td>70.8</td>
<td>72.9</td>
<td>73.7</td>
<td>74.2</td>
</tr>
</tbody>
</table>

(Quirk et al., 2015): IFTTT baselines
Example: what is the earliest flight from ci0 to ci1 tomorrow
Contributions

- Encoder-Decoder with Neural Networks
  - Seq2Seq/Seq2Tree models perform competitively on semantic parsing
  - A good task to understand the limitations of neural networks

Requirement of Captured Meaning

- Topic classification
- Sentiment classification
- Semantic parsing (complete meaning)
- ...
Contributions

- Encoder-Decoder with Neural Networks
  
  Seq2Seq/Seq2Tree models perform competitively on semantic parsing

- Tree decoder
  
  - Utilizing hierarchical structure of logical form improves performance
  
  - Structure prior/constraint of decoding results
  
  - Compositional nature of logical form
Contributions

- Encoder-Decoder with Neural Networks
  Seq2Seq/Seq2Tree models perform competitively on semantic parsing

- Tree decoder
  Utilizing hierarchical structure of logical form improves performance

- Attention mechanism
  Learn soft alignments between question and logical form
Future Work

- Weakly supervised learning
  Learn from (question, answer) pairs
- Open-domain
  QA over Freebase
- Utilize parsing results of questions
  CCG / Dependency / AMR
- Apply Seq2Tree model to related structured prediction tasks
Thanks!

Q&A

(Q-🙋→🤖-👨→A)

Code Available:
http://homepages.inf.ed.ac.uk/s1478528