Cross-lingual transfer of a semantic parser via parallel data

Kilian Evang
k.evang@rug.nl

University of Groningen

CLIN26. VU Amsterdam
2015-12-18
Introduction

Meaning Annotation by Proxy

Inducing Lexical Items Using Word Alignments

Shift-reduce Parsing

Experiments and Results
Semantic parsing: what?

From Words to (Logical) Meaning

*She likes to read books* →

DRT [Kamp, 1984]
Semantic parsing: why?

Translate to something a computer can “understand”

- commands for robots, e.g. [Dukes, 2014]
- queries for databases, e.g. [Reddy et al., 2014]
- formulas for (probabilistic) reasoners, e.g. [Beltagy et al., 2015]
Semantic parsing: how?

System for English [Curran et al., 2007]
System for other languages?
Learn (rudimentary) semantic parser from nothing but
- existing source language system (C&C+Boxer)
- parallel data
- (POS tagger for target language)
Method

1. meaning annotation by proxy
2. inducing lexical items using word alignments
3. shift-reduce parsing
Introduction

Meaning Annotation by Proxy

Inducing Lexical Items Using Word Alignments

Shift-reduce Parsing

Experiments and Results
Parallel corpus: Tatoeba.org

She likes to read books.

Sie liest gern Bücher.

Ši ťatas legi librojn.

Elle aime lire des livres.

Szeret könyvet olvasni.

Szeret könyveket olvasni.

Le piace leggere libri.

A lei piace leggere libri.

彼女は本を読むのが好きだ。

Ze leest graag boeken.

Ela gosta de ler livros.

Le gusta leer libros.

O kitap okumayı seviyor.
Automatic annotation of English sentences

<table>
<thead>
<tr>
<th>She</th>
<th>likes</th>
<th>to</th>
<th>read</th>
<th>books</th>
</tr>
</thead>
<tbody>
<tr>
<td>NP</td>
<td>(S[dcl] \ NP) / (S[to] \ NP)</td>
<td>(S[to] \ NP) / (S[b] \ NP)</td>
<td>(S[b] \ NP) / NP</td>
<td>NP</td>
</tr>
<tr>
<td>she'</td>
<td>like'</td>
<td>to'</td>
<td>read'@book'</td>
<td>book' &gt; 0</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>S[to] \ NP</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>to'@read'@book'</td>
<td>&gt; 0</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>S[to] \ NP</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>to'@read'@book'</td>
<td>&gt; 0</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>S[dcl] \ NP</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>like'@to'@read'@book'</td>
<td>&lt; 0</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>S[dcl]</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>like'@to'@read'@book')@she'</td>
<td></td>
</tr>
</tbody>
</table>
Meaning annotation by proxy

(like'@(to'@(read'@book'))))@she'
Ze leest graag boeken
Introduction

Meaning Annotation by Proxy

Inducing Lexical Items Using Word Alignments

Shift-reduce Parsing

Experiments and Results
Generating candidate lexical items

- [Zettlemoyer and Collins, 2007, Kwiatkowski et al., 2013]: hand-written lexical templates for English
- [Kwiatkowski et al., 2011]: recursively splitting gold-standard meaning representations, heuristics to constrain search space
- this work: from the English parse tree
  - use the same lexical semantics as in English
  - assign them to Dutch words, possibly one to two or two to one
  - drop category subdistinctions (dcl, b, to...)
  - use undirected slashes
Example alignment (ideal)

```
<table>
<thead>
<tr>
<th>NP</th>
<th>(S[dcl]\NP)/(S[to]\NP)</th>
<th>(S[to]\NP)/(S[b]\NP)</th>
<th>(S[b]\NP)/NP</th>
<th>NP</th>
</tr>
</thead>
<tbody>
<tr>
<td>she'</td>
<td>likes</td>
<td>to'</td>
<td>read'</td>
<td>book'</td>
</tr>
<tr>
<td>She</td>
<td>likes</td>
<td>to'</td>
<td>read</td>
<td>books</td>
</tr>
<tr>
<td>Ze</td>
<td>leest</td>
<td>graag</td>
<td>boeken</td>
<td></td>
</tr>
<tr>
<td>NP</td>
<td>(S</td>
<td>NP)</td>
<td>NP</td>
<td>(S</td>
</tr>
<tr>
<td>she'</td>
<td>read'</td>
<td>λx.(likes'@(to'@x))</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
```
Inducing Dutch lexical items

- extract one lexical item per translation unit, keep most frequent ones
- IBM model 4, all translation units from 5-best alignments in both directions
- for each word+POS, cutoff frequency is 0.1 of max
Introduction

Meaning Annotation by Proxy

Inducing Lexical Items Using Word Alignments

Shift-reduce Parsing

Experiments and Results
Shift-reduce parsing

- Based on English CCG parser of [Zhang and Clark, 2011]
- Action types: shift, combine, unary, skip, finish
- Allows fragmentary parses
Forced decoding

- We have:
  - 13,122 Dutch training sentences with target semantics
  - A CCG lexicon for Dutch
- We need:
  - Training parses for Dutch
- Solution: forced decoding with heuristic pruning based on target semantics [Zhao and Huang, 2015]
  - Training parses found for 4,038 sentences
  - Other 9,084: no parse found, or agenda explodes
Forced decoding

- We have:
  - 13,122 Dutch training sentences with target semantics
  - A CCG lexicon for Dutch
- We need:
  - Training parses for Dutch
- Solution: forced decoding with heuristic pruning based on target semantics [Zhao and Huang, 2015]
  - Training parses found for 4,038 sentences
  - Other 9,084: no parse found, or agenda explodes
Forced decoding

- We have:
  - 13,122 Dutch training sentences with target semantics
  - A CCG lexicon for Dutch
- We need:
  - Training parses for Dutch
- Solution: forced decoding with heuristic pruning based on target semantics [Zhao and Huang, 2015]
  - Training parses found for 4,038 sentences
  - Other 9,084: no parse found, or agenda explodes
Dutch training parse (example)

Ze

NP

leest

(S|NP)|NP

graag

λx.(likes′ @ (to′ @ x))

(S|NP)|(S|NP)

boeken

NP

book′

(S|NP)|NP

λx.(likes′ @ (to′ @ (read′ @ x)))

(S|NP)|NP

> 0

S

likes′ @ (to′ @ (read′ @ book′))

S

(likes′ @ (to′ @ (read′ @ book′))) @ book′

< 0
Parser training

• Training data: Dutch derivations obtained with forced decoding
• Averaged perceptron with beam search \((b = 16)\)
• Early update [Collins and Roark, 2004]
• Features: [Zhang and Clark, 2011]
Dealing with unknown words at test time

Pick schematic lexical entries for POS extracted from lexicon, e.g.

\[ f = \lambda x. \underline{\text{UNKNOWN}}(x) \]
Introduction

Meaning Annotation by Proxy

Inducing Lexical Items Using Word Alignments

Shift-reduce Parsing

Experiments and Results
Evaluation: graph match measure

[Allen et al., 2008, Le and Zuidema, 2012]
Evaluation: baseline and upper bound

- baseline: most frequent lexical entry/schema for each word, all unconnected
- upper bound: silver standard, unseen symbols replaced with \_\_UNKNOWN\_\_
Evaluation: baseline and upper bound

- baseline: most frequent lexical entry/schema for each word, all unconnected
- upper bound: silver standard, unseen symbols replaced with `__UNKNOWN__`
## Results on development test set (1,641 sentences)

<table>
<thead>
<tr>
<th></th>
<th>rec</th>
<th>prec</th>
<th>f1</th>
</tr>
</thead>
<tbody>
<tr>
<td>baseline</td>
<td>.338</td>
<td>.344</td>
<td>.341</td>
</tr>
<tr>
<td>training iterations:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>.362</td>
<td>.384</td>
<td>.372</td>
</tr>
<tr>
<td>1</td>
<td>.507</td>
<td>.503</td>
<td>.505</td>
</tr>
<tr>
<td>2</td>
<td>.504</td>
<td>.510</td>
<td>.507</td>
</tr>
<tr>
<td>3</td>
<td>.508</td>
<td>.514</td>
<td>.511</td>
</tr>
<tr>
<td>4</td>
<td>.510</td>
<td>.516</td>
<td>.513</td>
</tr>
<tr>
<td>5</td>
<td>.507</td>
<td>.512</td>
<td>.510</td>
</tr>
<tr>
<td>...</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>upper bound</td>
<td>.962</td>
<td>.896</td>
<td>0.928</td>
</tr>
</tbody>
</table>
Where it goes well
Introduction

Meaning Annotation

Inducing Lexical Items

Parsing

Results

Where it doesn’t
Conclusions

- CCG suitable formalism for cross-lingual semantic parser induction
- Reasonable grammar learned Dutch
- Important areas for future work
  - Lexicon induction: tweak to get more training data
  - Treat English parses, word alignments as latent
  - Morphology
  - Lexical semantics

Interested in collaborating? Let me know!
Conclusions

• CCG suitable formalism for cross-lingual semantic parser induction
• Reasonable grammar learned Dutch
• Important areas for future work
  • Lexicon induction: tweak to get more training data
  • Treat English parses, word alignments as latent
  • Morphology
  • Lexical semantics

Interested in collaborating? Let me know!
References I

Deep Semantic Analysis of Text.

Representing meaning with a combination of logical form and vectors.
CoRR, abs/1505.06816.

Incremental parsing with the perceptron algorithm.
In Proceedings of the 42nd Annual Meeting of the Association for Computational Linguistics (ACL-04).

Linguistically Motivated Large-Scale NLP with C&C and Boxer.

Semeval-2014 task 6: Supervised semantic parsing of robotic spatial commands.

A Theory of Truth and Semantic Representation.
References II

Scaling semantic parsers with on-the-fly ontology matching.

Lexical generalization in CCG grammar induction for semantic parsing.

Learning compositional semantics for open domain semantic parsing.

Large-scale semantic parsing without question-answer pairs.

Online learning of relaxed CCG grammars for parsing to logical form.

Shift-reduce CCG parsing.
Zhao, K. and Huang, L. (2015).
Type-driven incremental semantic parsing with polymorphism.
In Proceedings of NAACL HLT.