Weakly-Supervised Grammar-Informed Bayesian CCG Parser Learning

Dan Garrette  UT-Austin
Chris Dyer  CMU
Jason Baldridge  UT-Austin
Noah A. Smith  CMU
Motivation

Annotating parse trees \textit{by hand} is extremely difficult.
Motivation

Can we learn new parsers cheaply?

(cheaper = less supervision)
Motivation

When supervision is scarce, we have to be smarter about data.
Type-Level Supervision
Type-Level Supervision

- Unannotated text
- Incomplete tag dictionary: word $\mapsto$ \{tags\}
Type-Level Supervision

Used for part-of-speech tagging for 20+ years

[Kupiec, 1992]
[Merialdo, 1994]
Type-Level Supervision

Good tagger performance even with low supervision

[Ravi & Knight, 2009]
[Das & Petrov, 2011]
[Garrette & Baldridge, 2013]
[Garrette et al., 2013]
Combinatory Categorial Grammar (CCG)
CCG

Every word token is associated with a category.

Categories combine to form categories of larger constituents.

[Steedman, 2000]
[Steedman and Baldridge, 2011]
CCG

```
np
  np/n
  the

n
  dog
```
CCG

s

np    s\np

dogs  sleep
Type-Supervised CCG

<table>
<thead>
<tr>
<th></th>
<th>the</th>
<th>lazy</th>
<th>dogs</th>
<th>wander</th>
</tr>
</thead>
<tbody>
<tr>
<td>np/n</td>
<td>n/n</td>
<td>n</td>
<td>n</td>
<td>n</td>
</tr>
<tr>
<td>np</td>
<td>np</td>
<td>n/n</td>
<td>np/n</td>
<td>s\np</td>
</tr>
<tr>
<td>np/n</td>
<td>n/n</td>
<td>n</td>
<td>s\np</td>
<td></td>
</tr>
<tr>
<td>------</td>
<td>-----</td>
<td>---</td>
<td>------</td>
<td></td>
</tr>
<tr>
<td>the</td>
<td>lazy</td>
<td>dogs</td>
<td>wander</td>
<td></td>
</tr>
</tbody>
</table>
CCG Parsing

the lazy dogs wander

np/n  n/n  n  s\np
wander the lazy dogs
The lazy dogs wander.
CCG Parsing

[Diagram showing a syntax tree with nodes labeled as follows:
- np/n
- n/n
- n
- s\np

Words associated with each node include:
- the
- lazy
- dogs
- wander]
The phrase in the diagram is "the lazy dogs wander."
CCG Parsing

```
np
  / \  
np/n n/n n
  the lazy dogs
  s\np wander
```
CCG Parsing

Tree representation:

```
  np
   /\  
  n  n
 / \ / \  
np/n n/n n   s\np
    the lazy dogs wander
```
CCG Parsing

```
np
 /  
np/n  n/n  n  s\np

the  lazy  dogs  wander
```
Why CCG?

Machine Translation
[Weese, Callison-Burch, and Lopez, 2012]

Semantic Parsing
[Zettlemoyer and Collins, 2005]
Type-Supervised CCG

Type-supervised learning for CCG is highly *ambiguous*

**Penn Treebank parts-of-speech**
- 48 tags

**CCGBank Categories**
- 1,300+ categories
Our Strategy

The grammar formalism *itself* can be used to guide learning
Our Strategy

Incorporate *universal knowledge* about grammar into learning
Universal Knowledge
Prefer Simpler Categories

```
np
  \ni
  \np\(\np/\n\)\n    \ni
    np/\n
the    lazy    dog
```

```
np
  \ni
  \n\nnp/\n
the    lazy    dog
```
Prefer Simpler Categories

\[
\text{the} \quad \text{np}\backslash\text{(np/n)} \quad \text{lazy} \quad \text{dog}
\]

\[
\text{the} \quad \text{n/n} \quad \text{lazy} \quad \text{dog}
\]
Prefer Simpler Categories

buy := (s_b\np)/np \quad \text{appears 342 times in CCGbank}

e.g. “Opponents don't \textbf{buy} such arguments.”

buy := (((s_b\np)/np)/np)/np \quad \text{appears once}

“Tele-Communications agreed to \textbf{buy} half of Showtime Networks from Viacom for $225 million.”
Prefer Modifier Categories

\[
\text{(s_b\np)/np}
\]

transitive verb: (he) \textbf{hides} (the money)

\[
\text{(s_b\np)/np)/(s_b\np)/np}
\]

adverb: (he) \textbf{quickly} (hides) (the money)
Weighted Category Grammar

\[
\begin{align*}
a \quad \rightarrow & \quad \{s, np, n, \ldots\} & p_{\text{atom}}(a) \times p_{\text{term}} \\
A \quad \rightarrow & \quad B / B & p_{\text{term}} \times p_{\text{fwd}} \times p_{\text{mod}} \\
A \quad \rightarrow & \quad B / C & p_{\text{term}} \times p_{\text{fwd}} \times p_{\text{mod}} \\
A \quad \rightarrow & \quad B \backslash B & p_{\text{term}} \times p_{\text{fwd}} \times p_{\text{mod}} \\
A \quad \rightarrow & \quad B \backslash C & p_{\text{term}} \times p_{\text{fwd}} \times p_{\text{mod}}
\end{align*}
\]
Weighted Category Grammar

\[ a \rightarrow \{s, np, n, \ldots\} \]

\[ p_{\text{atom}}(a) \times p_{\text{term}} \]

\[ A \rightarrow B / B \]

\[ p_{\text{term}} \times p_{\text{fwd}} \times p_{\text{mod}} \]

\[ A \rightarrow B / C \]

\[ p_{\text{term}} \times p_{\text{fwd}} \times p_{\text{mod}} + p_{\text{term}} \times p_{\text{fwd}} \times p_{\text{mod}} \]

\[ A \rightarrow B \backslash B \]

\[ p_{\text{term}} \times p_{\text{fwd}} \times p_{\text{mod}} + p_{\text{term}} \times p_{\text{fwd}} \times p_{\text{mod}} \]

\[ A \rightarrow B \backslash C \]
Prefer Likely Categories

```
s
  np

    n

      np/n

    n/n

      the

    n

      lazy

    np

      dogs

      wander
```
Prefer Likely Categories

```
  s
 /  \
np  n
 /  \    /
np/n  n/n  n  s
  the   lazy  dogs   wander
```
Type-Supervised Learning

unlabeled corpus

unlabeled corpus

tag dictionary

universal properties of the CCG formalism

same as POS tagging
Posterior Inference

[Johnson, Griffiths, and Goldwater, 2007]
Posterior Inference

Priors (simple is good)

PCFG

Inside

the (np/n) lazy (n/n) dogs (n) wander (n)

(np) (n) (np) (s\np) (n/n) (n/n) (s\np) ...

Inside
Posterior Inference

Priors (simple is good)

PCFG

Inside
Posterior Inference

Priors (simple is good)

PCFG

the
np/n

lazy
n/n
np

dogs
n
np

(s\np)/np

wander
n
n/n

np/n

s\np

...
Posterior Inference

Priors (simple is good)

PCFG

Sample

the np/n

lazy n/n

np

dogs n

np

(s\np)/np

wander n

n/n

np/n

...
Posterior Inference

Priors (simple is good)

PCFG

the
np/n

lazy
n/n np

dogs
n np
(np)/np

wander
n n/n np
(s\np)/n

s\np

...
Posterior Inference

Priors (simple is good)

PCFG

the
np/n

lazy
n/n
np

dogs
n
np
(np)/np

wander
n
n/n
np/n
...
Posterior Inference

Priors (simple is good)

PCFG

the (np/n)
lazy (n/n)
dogs (n)
wander (n/n)

(s\np)/np

np

np

...
Results
CCG Parsing Results

<table>
<thead>
<tr>
<th>Language</th>
<th>Uniform</th>
<th>With Prior</th>
</tr>
</thead>
<tbody>
<tr>
<td>English</td>
<td>53.4</td>
<td>55.7</td>
</tr>
<tr>
<td>Chinese</td>
<td>35.9</td>
<td>42.0</td>
</tr>
<tr>
<td>Italian</td>
<td>58.2</td>
<td>60.0</td>
</tr>
</tbody>
</table>
Conclusion

Using *universal grammatical knowledge* can make better use of weak supervision