Adaptive Supertagging for Faster Parsing

Jonathan K. Kummerfeld
Supervisor – Dr. James Curran
Motivation – Natural Language Processing

Build systems that use text intelligently

- Question Answering
  "When was the PM of Australia not the leader of the majority party in the House of Representatives?"

- Machine Translation
  Translate this website from French into English

- Anaphora Resolution
  "Steve is the leader of Apple, and without him it would be far less successful."

All of these rely on syntactic information, produced by parsers
Tagging and Parsing

One claims he is pro-choice
Part of Speech Tagging

One claims he is pro-choice

NN VBZ PRP VBZ JJ
Combinatory Categorial Grammar (CCG) – Supertagging

One claims he is pro-choice

\[
\begin{align*}
N &\quad (S\backslash NP)/S \\
NP &\quad (S\backslash NP)/(S\backslash NP) \\
S\backslash NP &\quad S\backslash NP
\end{align*}
\]
Combinatory Categorial Grammar (CCG) – Parsing

One claims he is pro–choice

\[
\begin{align*}
N & \quad (S\backslash NP)/S \\
NP & \quad (S\backslash NP)/(S\backslash NP) \\
NP & \quad (S\backslash NP)/(S\backslash NP) \\
S\backslash NP & \quad \text{pro–choice}
\end{align*}
\]
Motivation – Parsing

Parsers are slow:

- State-of-the-art, usually <1 sentence / sec
- Fastest state-of-the-art, <50 sentences / sec

Especially for our data sets:

- Estimated 10,000,000,000,000 words online
- More coming
The key idea behind the speed of the fastest parsers today is to shift work from parsing to tagging:
For $n$ words, each with $k$ tags

- Tagging – $O(nk)$
- Parsing – $O(n^3k^2)$
Contributions

Core Novel Ideas

- Adapt the supertagger to provide the tags the parser will use
- Parser performance is not proportional to supertagging accuracy

Development

- Perceptron Algorithms
- Parallelisation

Demonstration

- Improved performance
- Domain adaptation
- Explored Supertagger – Parser interaction
Ideal World

One claims he is pro-choice

\[
\begin{align*}
N & \quad (S\backslash NP)/S \\
NP & \quad (S\backslash NP)/(S\backslash NP) \\
S\backslash NP & \quad S\backslash NP
\end{align*}
\]
Current World – Problem

One claims he is pro–choice.

\[
\begin{array}{c}
N \\
(S\backslash NP)/NP
\end{array}
\quad
\begin{array}{c}
NP \\
(S\backslash NP)/(S\backslash NP)
\end{array}
\quad
\begin{array}{c}
S\backslash NP
\end{array}
\]
Current World – Solution

<table>
<thead>
<tr>
<th>One</th>
<th>claims</th>
<th>he</th>
<th>is</th>
<th>pro – choice</th>
</tr>
</thead>
<tbody>
<tr>
<td>(N/N)</td>
<td>((S\backslash NP)/NP)</td>
<td>(NP)</td>
<td>((S\backslash NP)/(S\backslash NP))</td>
<td>(S\backslash NP)</td>
</tr>
<tr>
<td>(N)</td>
<td>(N)</td>
<td>(NP)</td>
<td>((S\backslash NP)/NP)</td>
<td>((S\backslash NP)/(S\backslash NP))</td>
</tr>
<tr>
<td>((S/S)/(S/S))</td>
<td>((S\backslash NP)/(S\backslash NP))</td>
<td>((S\backslash NP)/(S\backslash NP))</td>
<td>((S\backslash NP)/(S\backslash NP))</td>
<td>((S\backslash NP)/(S\backslash NP))</td>
</tr>
</tbody>
</table>

Jonathan K. Kummerfeld
Adaptive Supertagging for Faster Parsing
Adaptive Supertagging

One claims he is pro-choice

How do we teach the supertagger to produce these tags?
Use the parser!
Contributions

Core Novel Ideas

- Adapt the supertagger to provide the tags the parser will use
- Parser performance is not proportional to supertagging accuracy

Development

- Perceptron Algorithms
- Parallelisation

Demonstration

- Improved performance
- Domain adaptation
- Explored Supertagger – Parser interaction
## Implementation

<table>
<thead>
<tr>
<th>Component</th>
<th>Initial System</th>
<th>Additions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Statistical Feature Extraction</td>
<td>3 Types</td>
<td>+9 Types</td>
</tr>
<tr>
<td></td>
<td>Single thread</td>
<td>Parallel</td>
</tr>
<tr>
<td>Parameter Estimation</td>
<td>BFGS, GIS</td>
<td>AP, MIRA</td>
</tr>
<tr>
<td></td>
<td>Single thread</td>
<td>Parallel</td>
</tr>
</tbody>
</table>
Implementation – Initial System

1. data
2. extract features
3. contexts
4. estimate weights
5. model
Implementation – Parallelised

Jonathan K. Kummerfeld
Adaptive Supertagging for Faster Parsing
Implementation – Parallelised Weight Estimation
## Implementation – AP and MIRA

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>Training Time (sec)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>40k</td>
</tr>
<tr>
<td>GIS</td>
<td>7,200</td>
</tr>
<tr>
<td>BFGS</td>
<td>6,300</td>
</tr>
<tr>
<td>AP</td>
<td>76</td>
</tr>
<tr>
<td>MIRA</td>
<td>96</td>
</tr>
</tbody>
</table>
Contributions

Core Novel Ideas

- Adapt the supertagger to provide the tags the parser will use
- Parser performance is not proportional to supertagging accuracy

Development

- Perceptron Algorithms
- Parallelisation

Demonstration

- Improved performance
- Domain adaptation
- Exploration of Supertagger – Parser interaction
Tests

> 300 Models trained
  
  - Algorithm – GIS, BFGS, AP, MIRA
  - Various amounts of training data – 40k, 80k ... 26,000k
  - Number of cores – 1, 2 ... 64
  - Statistical features – Subtractive analysis of 9 new feature types

> 12,000 Tests

  - Tested with tuned and default settings
  - Evaluated on Wall Street Journal and Wikipedia
  - Systematic analysis of 1,913 sentences parsed with 2,200 variations of parameters
Improved Performance

**Figure:** Evaluation on the Wall Street Journal

Jonathan K. Kummerfeld

Adaptive Supertagging for Faster Parsing
Improved Performance

Figure: Evaluation on the Wall Street Journal
Domain Adaptation

Figure: Evaluation on Wikipedia

Jonathan K. Kummerfeld  Adaptive Supertagging for Faster Parsing
Domain Adaptation

Figure: Evaluation on Wikipedia

Jonathan K. Kummerfeld
Adaptive Supertagging for Faster Parsing
Domain Adaptation

Figure: Evaluation on Wikipedia
## Contributions

<table>
<thead>
<tr>
<th>Metric</th>
<th>Baseline</th>
<th>Final Models</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Wall Street Journal</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Accuracy (%)</td>
<td>83.41</td>
<td>83.99</td>
</tr>
<tr>
<td>Speed (sents / sec)</td>
<td>48.5</td>
<td>90.2</td>
</tr>
<tr>
<td><strong>Wikipedia</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Accuracy (%)</td>
<td>82.5</td>
<td>83.3</td>
</tr>
<tr>
<td>Speed (sents / sec)</td>
<td>46.3</td>
<td>60.5</td>
</tr>
</tbody>
</table>

This will lead directly to benefits in NLP systems.
Acknowledgements

- Johns Hopkins University, CLSP Summer Workshop in particular, Jessika Rosener, collaborator on parallelisation
- Early results to appear as:
  Jonathan K. Kummerfeld and Dr James R. Curran Faster Parsing and Supertagging Model Estimation, ALTA 2009
Second Core Idea

**Figure:** Parsing behaviour for the 1791\textsuperscript{st} sentence in section 00.