Event Extraction as Dependency Parsing
(in BioNLP 2011)

David McClosky
Stanford University
6.24.2011

Joint work with Mihai Surdeanu and Christopher D. Manning
Our approach in two slides...

... tax, acts as a costimulatory signal for GM-CSF and IL-2 gene transcription...
Our approach in two slides...

-root-

Full details in [McClosky, Surdeanu, and Manning, ACL 2011]
Outline

1. Event Parsing
2. Adapting to BioNLP 2011
3. Experiments
4. Conclusion
... tax, acts as a costimulatory signal for GM-CSF and IL-2 gene transcription ...
Approach

Preprocessing: Segmentation, tokenization, syntactic parsing

Self-trained biomedical parser: [McClosky, 2010]

... tax, acts as a costimulatory signal for GM-CSF and IL-2 gene transcription ...
Anchor classification: Token classification for event anchors

(similar to [Björne et al., BioNLP 2009])
Event parsing: Parse anchors and proteins using reranking parser
Maximum-spanning tree based parsing

Why a dependency parser?

- Event structures are non-projective (non-planar)
Maximum-spanning tree based parsing

Why a dependency parser?
- Event structures are non-projective (non-planar)

Why MSTParser? [McDonald et al., EMNLP 2005]
- Handles non-projective trees naturally
Maximum-spanning tree based parsing

Why a dependency parser?
- Event structures are non-projective (non-planar)

Why MSTParser? [McDonald et al., EMNLP 2005]
- Handles non-projective trees naturally
- Easy to extend feature extractor
Maximum-spanning tree based parsing

Why a dependency parser?
- Event structures are non-projective (non-planar)

Why MSTParser? [McDonald et al., EMNLP 2005]
- Handles non-projective trees naturally
- Easy to extend feature extractor
- Support for \( n \)-best parsing
Adapting to BioNLP 2011

- General improvements
  - Distributional similarity features in anchor detection
Adapting to BioNLP 2011

General improvements

- Distributional similarity features in anchor detection
- Improved head percolation rules for multiword anchors
Adapting to BioNLP 2011

General improvements
- Distributional similarity features in anchor detection
- Improved head percolation rules for multiword anchors
- Using lemmas (along with word forms) during event parsing
General improvements
- Distributional similarity features in anchor detection
- Improved head percolation rules for multiword anchors
- Using lemmas (along with word forms) during event parsing

Domain-specific customization
- Update event type information (EPI, ID)
Adapting to BioNLP 2011

- General improvements
  - Distributional similarity features in anchor detection
  - Improved head percolation rules for multiword anchors
  - Using lemmas (along with word forms) during event parsing

- Domain-specific customization
  - Update event type information (EPI, ID)
  - Combine ID training data with GENIA (ID)
Adapting to BioNLP 2011

General improvements
- Distributional similarity features in anchor detection
- Improved head percolation rules for multiword anchors
- Using lemmas (along with word forms) during event parsing

Domain-specific customization
- Update event type information (EPI, ID)
- Combine ID training data with GENIA (ID)
- Removing nested entities (ID)
## Results on Genia development

<table>
<thead>
<tr>
<th>Decoder(s)</th>
<th>Parser</th>
<th>Reranker</th>
</tr>
</thead>
<tbody>
<tr>
<td>1P</td>
<td>49.0</td>
<td>49.4</td>
</tr>
<tr>
<td>2P</td>
<td>49.5</td>
<td>50.5</td>
</tr>
<tr>
<td>1N</td>
<td>49.9</td>
<td>50.2</td>
</tr>
<tr>
<td>2N</td>
<td>46.5</td>
<td>47.9</td>
</tr>
<tr>
<td><strong>All</strong></td>
<td>—</td>
<td><strong>50.7</strong></td>
</tr>
</tbody>
</table>
Results on Epigenetics development

<table>
<thead>
<tr>
<th>Decoder(s)</th>
<th>Parser</th>
<th>Reranker</th>
</tr>
</thead>
<tbody>
<tr>
<td>1P</td>
<td>62.3</td>
<td>63.3</td>
</tr>
<tr>
<td>2P</td>
<td>62.2</td>
<td>63.3</td>
</tr>
<tr>
<td>1N</td>
<td>62.9</td>
<td>64.6</td>
</tr>
<tr>
<td>2N</td>
<td>60.8</td>
<td>63.8</td>
</tr>
<tr>
<td>All</td>
<td>—</td>
<td>64.1</td>
</tr>
</tbody>
</table>

(note: issues with our internal evaluator implementation)
## Domain adaptation for Infectious Diseases

<table>
<thead>
<tr>
<th>Model</th>
<th>Precision</th>
<th>Recall</th>
<th>f-score</th>
</tr>
</thead>
<tbody>
<tr>
<td>ID</td>
<td>59.3</td>
<td>38.0</td>
<td>46.3</td>
</tr>
<tr>
<td>ID (×1) + GE</td>
<td>52.0</td>
<td>40.2</td>
<td>45.3</td>
</tr>
<tr>
<td>ID (×2) + GE</td>
<td>52.4</td>
<td>41.7</td>
<td>46.4</td>
</tr>
<tr>
<td>ID (×3) + GE</td>
<td>54.8</td>
<td>45.0</td>
<td>49.4</td>
</tr>
<tr>
<td>ID (×4) + GE</td>
<td>55.2</td>
<td>43.8</td>
<td>48.9</td>
</tr>
<tr>
<td>ID (×5) + GE</td>
<td>55.1</td>
<td>44.7</td>
<td>49.4</td>
</tr>
</tbody>
</table>

(parser only with 2N decoder)
## Results on Infectious Diseases development

<table>
<thead>
<tr>
<th>Decoder(s)</th>
<th>Parser</th>
<th>Reranker</th>
</tr>
</thead>
<tbody>
<tr>
<td>1P</td>
<td>46.0</td>
<td>48.5</td>
</tr>
<tr>
<td>2P</td>
<td>47.8</td>
<td>49.8</td>
</tr>
<tr>
<td>1N</td>
<td>48.5</td>
<td>49.4</td>
</tr>
<tr>
<td>2N</td>
<td><strong>49.4</strong></td>
<td>48.8</td>
</tr>
<tr>
<td>All</td>
<td>—</td>
<td><strong>50.2</strong></td>
</tr>
</tbody>
</table>

---

David McClosky (Stanford)  
Event Parsing in BioNLP 2011  
6.24.2011 9 / 10
Summary

- New approach to event extraction
  - Parsing can be used for event extraction
  - Reranker further improves performance
- Minimal changes to adapt to new BioNLP domains
- Component in the FAUST system (stay tuned!)
- Code coming soon!
  
Summary

- New approach to event extraction
  - Parsing can be used for event extraction
  - Reranker further improves performance
- Minimal changes to adapt to new BioNLP domains
- Component in the FAUST system (stay tuned!)
- Code coming soon!

http://nlp.stanford.edu/software/eventparsing.shtml

Questions?