Neural CRF Parsing

Greg Durrett and Dan Klein
UC Berkeley
Parsing with CKY
He gave a long speech on foreign policy.
He gave a long speech on foreign policy.
He gave a long speech on foreign policy.
Parsing with CKY

He gave a long speech on foreign policy.
He gave a long speech on foreign policy.
Discrete Structure

He gave a long speech on foreign policy.
He gave a long speech on foreign policy.
He gave a long speech on foreign policy.
He gave a long speech on foreign policy.

[Socher et al. (2013)]
He gave a long speech on foreign policy.

[Socher et al. (2013)]
He gave a long speech on foreign policy.

Powerful nonlinear featurization...

[Socher et al. (2013)]
He gave a long speech on foreign policy.

Powerful nonlinear featurization...but inference is now intractable.

[Socher et al. (2013)]
CRF Parsing

He gave a long speech on foreign policy.

[Taskar et al. (2004), Finkel et al. (2008), Hall et al. (2014)]
He gave a long speech on foreign policy.

[Taskar et al. (2004), Finkel et al. (2008), Hall et al. (2014)]
He gave a long speech on foreign policy.

[Taskar et al. (2004), Finkel et al. (2008), Hall et al. (2014)]
CRF Parsing

He gave a long speech on foreign policy.

[Taskar et al. (2004), Finkel et al. (2008), Hall et al. (2014)]
CRF Parsing

He gave a long speech on foreign policy.

[Taskar et al. (2004), Finkel et al. (2008), Hall et al. (2014)]
CRF Parsing

He gave a long speech on foreign policy.

- Discrete structure with discrete features

[Taskar et al. (2004), Finkel et al. (2008), Hall et al. (2014)]
CRF Parsing

He gave a long speech on foreign policy.

- Discrete structure with discrete features
- Efficient inference via basic CKY...

[Taskar et al. (2004), Finkel et al. (2008), Hall et al. (2014)]
He gave a long speech on foreign policy.

- Discrete structure with discrete features
- Efficient inference via basic CKY... but the model is typically linear

[Taskar et al. (2004), Finkel et al. (2008), Hall et al. (2014)]
He gave a long speech on foreign policy.
Neural CRF Parsing

He gave a long speech on foreign policy.

Score:

a speech on policy
He gave a long speech on foreign policy.
He gave a long speech on foreign policy.
Neural CRF Parsing

He gave a long speech on foreign policy.
He gave a long speech on foreign policy.
Neural CRF Parsing

He gave a long speech on foreign policy.

- Neural networks score decisions locally (Collobert et al., 2011)
Neural networks score decisions locally (Collobert et al., 2011)

- Discrete structure with *continuous* features
Neural CRF Parsing

He gave a long speech on foreign policy.

- Neural networks score decisions locally (Collobert et al., 2011)
- Discrete structure with *continuous* features
- Inference is still efficient...
Neural CRF Parsing

He gave a long speech on foreign policy.

- Neural networks score decisions locally (Collobert et al., 2011)
- Discrete structure with continuous features
- Inference is still efficient...and we get nonlinear featurization!
Model
He gave a long speech on foreign policy.

[Hall, Durrett, Klein (2014)]
Basic CRF Model

\[ P(T|x) \propto \prod_{r \in T} \exp(\text{score}(r)) \]

He gave a long speech on foreign policy.

[Hall, Durrett, Klein (2014)]
Basic CRF Model

\[ P(T|x) \propto \prod_{r \in T} \exp(\text{score}(r)) \]

\[ \text{score}\left( \begin{array}{c}
\text{NP} \\
\text{NP} \\
\text{PP} \\
\end{array} \right) \]

He gave a long speech on foreign policy.

[Hall, Durrett, Klein (2014)]
Basic CRF Model

\[ P(T|x) \propto \prod_{r \in T} \exp(\text{score}(r)) \quad \text{score}(\begin{array}{c} \text{NP} \\ 2 \\ \text{NP} \\ 5 \\ \text{PP} \\ 8 \end{array}) = w^\top f(\begin{array}{c} \text{NP} \\ 2 \\ \text{NP} \\ 5 \\ \text{PP} \\ 8 \end{array}) \]

He gave a long speech on foreign policy.

[Hall, Durrett, Klein (2014)]
Basic CRF Model

\[ P(T|x) \propto \prod_{r \in T} \exp(\text{score}(r)) \]

\[ \text{score}\left( \begin{array}{c}
\text{NP} \\
2 \\
\text{NP} \\
5 \\
\text{PP} \\
8 \\
\end{array}\right) = w^\top f\left( \begin{array}{c}
\text{NP} \\
2 \\
\text{NP} \\
5 \\
\text{PP} \\
8 \\
\end{array}\right) \]

He gave a long speech on foreign policy.

[Hall, Durrett, Klein (2014)]
FirstWord = a ∧ NP

He gave a long speech on foreign policy.

[Hall, Durrett, Klein (2014)]
Basic CRF Model

\[ P(T|x) \propto \prod_{r \in T} \exp(\text{score}(r)) \]

\[ \text{score}\left(\begin{array}{c}
\text{NP} \\
\text{NP} \\
\text{PP} \\
\text{NP} \\
\text{PP} \\
\text{NP} \\
\text{PP} \\
\end{array}\right) = w^\top f\left(\begin{array}{c}
\text{NP} \\
\text{NP} \\
\text{PP} \\
\end{array}\right) \]

FirstWord = a \land

He gave a long speech on foreign policy.

[Hall, Durrett, Klein (2014)]
Basic CRF Model

\[ P(T|x) \propto \prod_{r \in T} \exp(\text{score}(r)) \]

\[ \text{score}\left(\begin{array}{c}
\text{NP} \\
\text{NP} \\
\text{PP} \\
\text{PP}
\end{array}\right) = w^\top f\left(\begin{array}{c}
\text{NP} \\
\text{NP} \\
\text{PP} \\
\text{PP}
\end{array}\right) \]

FirstWord = a \land 
\begin{array}{c}
\text{NP} \\
\text{PP}
\end{array}

PrevWord = gave \land 
\begin{array}{c}
\text{NP} \\
\text{PP}
\end{array}

He gave a long speech on foreign policy.

[Hall, Durrett, Klein (2014)]
Basic CRF Model

\[ P(T|x) \propto \prod_{r \in T} \exp(\text{score}(r)) \]

\[ \text{score} \left( \begin{array}{c} \text{NP} \\ 2 \\
\text{NP} \\ 5 \\
\text{PP} \\ 8 \end{array} \right) = w^\top f \left( \begin{array}{c} \text{NP} \\ 2 \\
\text{NP} \\ 5 \\
\text{PP} \\ 8 \end{array} \right) \]

FirstWord = a \land \begin{array}{c} \text{NP} \\ \text{PP} \\
\text{NP} \\ \text{PP} \end{array}

PrevWord = gave \land \begin{array}{c} \text{NP} \\ \text{PP} \\
\text{NP} \\ \text{PP} \end{array}

He gave a long speech on foreign policy.

[Hall, Durrett, Klein (2014)]
Basic CRF Model

\[
\text{score}\left( \begin{array}{c}
\text{NP} \\
2 \\
\text{NP} \\
5 \\
\text{PP} \\
8 \\
\end{array}\right) = w^\top f\left( \begin{array}{c}
\text{NP} \\
2 \\
\text{NP} \\
5 \\
\text{PP} \\
8 \\
\end{array}\right)
\]

FirstWord = a \land \begin{array}{c}
\text{NP} \\
\text{NP} \\
\text{PP} \\
\end{array}

PrevWord = gave \land \begin{array}{c}
\text{NP} \\
\text{NP} \\
\text{PP} \\
\end{array}
Basic CRF Model

\[
score\left(\begin{array}{c}
NP \\
2
\end{array}\right) = w^\top f\left(\begin{array}{c}
NP \\
5
\end{array}\right)
\]

Surface feature

FirstWord = a  \land  \begin{array}{c}
NP \\
2
\end{array}
\land
\begin{array}{c}
NP \\
5
\end{array}
\land
\begin{array}{c}
NP \\
8
\end{array}

PrevWord = gave  \land  \begin{array}{c}
NP \\
2
\end{array}
\land
\begin{array}{c}
NP \\
5
\end{array}
\land
\begin{array}{c}
NP \\
8
\end{array}
Basic CRF Model

\[
\text{score}\left(\left(\begin{array}{c}
\text{NP} \\
\text{NP} \\
\text{PP} \\
\end{array}\right)_{2,5,8}\right) = w^\top f\left(\left(\begin{array}{c}
\text{NP} \\
\text{NP} \\
\text{PP} \\
\end{array}\right)_{2,5,8}\right)
\]

Surface feature \hspace{1cm} Label feature

FirstWord = a \hspace{1cm} \land \hspace{1cm} \text{NP} \hspace{1cm} \text{NP} \hspace{1cm} \text{PP} \\
PrevWord = gave \hspace{1cm} \land \hspace{1cm} \text{NP} \hspace{1cm} \text{NP} \hspace{1cm} \text{PP}
**Basic CRF Model**

\[
\text{score}\left(\begin{array}{c}
\text{NP} \\
\text{NP}_2 \\
\text{NP}_5 \\
\text{PP} \\
\text{NP}
\end{array}\right) = w^\top f\left(\begin{array}{c}
\text{NP} \\
\text{NP}_2 \\
\text{NP}_5 \\
\text{PP} \\
\text{NP}
\end{array}\right)
\]

**Surface feature**

FirstWord = a

PrevWord = gave

**Label feature**

First = a, Prev = gave, ...

\( s \quad \bullet \cdots \bullet \cdots \bullet \quad \bullet \cdots \bullet \cdots \bullet \)

\( w^\top \)
Basic CRF Model

\[
score\left(\left( \begin{array}{c} \text{NP} \\ \text{NP} & \text{PP} \\ \text{NP} & \text{PP} & \text{NP} \\ \text{PP} & \text{NP} & \text{PP} \\ \text{NP} & \text{PP} & \text{NP} \\ \text{PP} & \text{NP} & \text{PP} \end{array} \right) \right) = w^\top f\left(\left( \begin{array}{c} \text{NP} \\ \text{NP} & \text{PP} \\ \text{NP} & \text{PP} & \text{NP} \\ \text{PP} & \text{NP} & \text{PP} \\ \text{NP} & \text{PP} & \text{NP} \\ \text{PP} & \text{NP} & \text{PP} \end{array} \right) \right)
\]

Surface feature

FirstWord = a

PrevWord = gave

Label feature

First = a, Prev = gave, ...
Basic CRF Model

\[
\text{score}\left(\begin{array}{c}
\text{NP} \\
\text{NP} \\
\text{PP} \\
\end{array}\right) = w^\top f\left(\begin{array}{c}
\text{NP} \\
\text{NP} \\
\text{PP} \\
\end{array}\right)
\]

Surface feature

FirstWord = a \land
PrevWord = gave \land

Label feature

\[F_{i,j} = s_i \ell_j\]

First = a, Prev = gave, ...
Basic CRF Model

\[
\text{score}\left(\begin{array}{c}
\text{NP} \\
\text{NP} \\
\text{PP} \\
\end{array}\right) = w^\top f\left(\begin{array}{c}
\text{NP} \\
\text{NP} \\
\text{PP} \\
\end{array}\right) = W \circ
\]

Surface feature
FirstWord = a
PrevWord = gave

Label feature
\(F_{i,j} = s_i \ell_j\)

First = a, Prev = gave, ...
Basic CRF Model

\[
\text{score} \left( \begin{array}{c}
\text{NP} \\
\text{NP} \\
\text{PP}
\end{array} \right) = w^\top f \left( \begin{array}{c}
\text{NP} \\
\text{NP} \\
\text{PP}
\end{array} \right) = W \odot s \left( \begin{array}{c}
x \\
x \\
x
\end{array} \right)
\]

Surface feature

FirstWord = a

PrevWord = gave

Label feature

\[ F_{i,j} = s_i \ell_j \]

First = a, Prev = gave, ...

\[
s(\cdot, \cdot, \cdot, \cdot, \cdot, \cdot, \cdot, \cdot, \cdot, \cdot, \cdot, \cdot, \cdot, \cdot)
\]
Basic CRF Model

\[
\text{score}\left(\begin{array}{c} \text{NP} \\ \text{NP} \\ \text{PP} \\ 2 \\ 5 \\ 8 \end{array}\right) = w^\top f\left(\begin{array}{c} \text{NP} \\ \text{NP} \\ \text{PP} \\ 2 \\ 5 \\ 8 \end{array}\right) = W \odot s \left(\begin{array}{c} X \\ X \\ X \\ 2 \\ 5 \\ 8 \end{array}\right) \ell^\top \left(\begin{array}{c} \text{NP} \\ \text{PP} \\ \text{NP} \end{array}\right)
\]

Surface feature

FirstWord = a \land \quad \text{PrevWord} = \text{gave} \land

Label feature

\[
F_{i,j} = s_i \ell_j
\]

First = a, Prev = gave, ...
Neural CRF Model

\[
score \left( \begin{array}{c}
\text{NP} \\
\text{NP} \\
\text{PP} \\
\end{array} \right) = W \odot s \left( \begin{array}{c}
\text{X} \\
\text{X} \\
\text{X} \\
\end{array} \right) \ell^T \left( \begin{array}{c}
\text{NP} \\
\text{NP} \\
\text{PP} \\
\end{array} \right)
\]
Neural CRF Model

\[
\text{score} \left( \begin{array}{c} \text{NP} \\
\text{NP} \\
\text{PP} \\
\end{array} \right) = W \odot s \left( \begin{array}{c} X \\
X \\
X \\
\end{array} \right) \ell^T \left( \begin{array}{c} \text{NP} \\
\text{NP} \\
\text{PP} \\
\end{array} \right)
\]
He gave a long speech on foreign policy.
He gave a long speech on foreign policy.
Neural CRF Model

\[
\text{score}(\begin{array}{c}
  2 \\
  \text{NP} \\
  5 \\
  \text{PP} \\
  8 
\end{array}) = W \odot s \left( \begin{array}{c}
  2 \\
  \text{X} \\
  5 \\
  \text{X} \\
  8 
\end{array} \right) \ell^T \left( \begin{array}{c}
  \text{NP} \\
  \text{PP} 
\end{array} \right)
\]

He gave a long speech on foreign policy.
He gave a long speech on foreign policy.
Neural CRF Model

\[
\text{score} \left( \begin{array}{c} \text{NP} \\ \text{NP} \\ \text{PP} \end{array} \right) = W \odot s \left( \begin{array}{c} X \\ X \\ X \\ X \end{array} \right) \ell^T \left( \begin{array}{c} \text{NP} \\ \text{PP} \end{array} \right)
\]

He gave a long speech on foreign policy.

100-dim vectors (Bansal et al., 2014)
He gave a long speech on foreign policy.
He gave a long speech on foreign policy.
Neural CRF Model

\[
score \left( \frac{\text{NP}}{2} \frac{\text{NP}}{5} \frac{\text{PP}}{8} \right) = W \odot s \left( \frac{\text{X}}{2} \frac{\text{X}}{5} \frac{\text{X}}{8} \right) \ell^T \left( \frac{\text{NP}}{2} \frac{\text{NP}}{5} \frac{\text{PP}}{8} \right)
\]

He gave a long speech on foreign policy.
Neural CRF Model

\[
\text{score}\left(\begin{array}{c}
\text{NP} \\
\text{NP} \\
\text{PP}
\end{array}\right) = W \odot s\left(\begin{array}{c}
\text{X} \\
\text{X} \\
\text{X}
\end{array}\right) \ell^T\left(\begin{array}{c}
\text{NP} \\
\text{NP} \\
\text{PP}
\end{array}\right)
\]

He gave a long speech on foreign policy.
Neural CRF Model

\[
\text{score} \left( \begin{array}{c}
\text{NP} \\
\text{NP} \\
\text{PP}
\end{array} \right) = W \odot s \left( \begin{array}{c}
\text{X} \\
\text{X} \\
\text{X}
\end{array} \right) \ell^T \left( \begin{array}{c}
\text{NP} \\
\text{NP} \\
\text{PP}
\end{array} \right)
\]

He gave a long speech on foreign policy.
Neural CRF Model

\[
\text{score} \left( \begin{array}{c}
\text{NP} \\
\text{NP} \\
\text{PP}
\end{array} \right) = W \odot s \left( \begin{array}{c}
\text{X} \\
\text{X} \\
\text{X}
\end{array} \right) l^T \left( \begin{array}{c}
\text{NP} \\
\text{NP} \\
\text{PP}
\end{array} \right)
\]
He gave a long speech on foreign policy.
Inference
Inference

Just CKY!
Inference

Just CKY!

... with coarse pruning and caching of neural net operations

(Goodman, 1997)  (Chen and Manning, 2014)
Inference

Just CKY!

... with coarse pruning and caching of neural net operations
(Goodman, 1997)  (Chen and Manning, 2014)

Roughly 2x slower than with sparse features alone
Learning
Learning

Just Maximum Likelihood!
Learning

Just Maximum Likelihood!

... with backpropagation through each local neural network
Optimization: Adadelta (Zeiler, 2012) worked slightly better than Adagrad (Duchi et al., 2011)
Results
Results: English Treebank (Dev)

Dev set $F_1$ all lengths

90.1
Sparse
Results: English Treebank (Dev)

Dev set F₁ all lengths

- Sparse: 90.1
- Neural: 90.4
Results: English Treebank (Dev)

Dev set F₁ all lengths

- Sparse: 90.1
- Neural: 90.4
- Sparse+: 91.3
Results: English Treebank (Dev)

<table>
<thead>
<tr>
<th>Method</th>
<th>Dev set F₁ all lengths</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sparse</td>
<td>90.1</td>
</tr>
<tr>
<td>Neural</td>
<td>90.4</td>
</tr>
<tr>
<td>Sparse+ Neural</td>
<td>91.3</td>
</tr>
<tr>
<td>Sparse+ Brown</td>
<td>90.2</td>
</tr>
</tbody>
</table>
Results: English Treebank (Dev)

Dev set $F_1$ all lengths

- Sparse: 90.1
- Sparse+ Neural: 91.3
- Sparse+ Brown: 90.2
Word Vectors

Dev set F1 all lengths

92
91
90
89
88
87
Word Vectors

Dev set F1 all lengths

- Bansal et al.: 90.4
- Collobert and Weston: 89.6
Word Vectors

Dependency context

Dev set F₁ all lengths

Bansal et al.

Collobert and Weston

11-word surface context

90.4

89.6
Syntactic vectors are best for parsing (Bansal et al., 2014; Levy and Goldberg, 2014)
Syntactic vectors are best for parsing (Bansal et al., 2014; Levy and Goldberg, 2014)
- Syntactic vectors are best for parsing (Bansal et al., 2014; Levy and Goldberg, 2014)
Syntactic vectors are best for parsing (Bansal et al., 2014; Levy and Goldberg, 2014)
Syntactic vectors are best for parsing (Bansal et al., 2014; Levy and Goldberg, 2014)
Syntactic vectors are best for parsing (Bansal et al., 2014; Levy and Goldberg, 2014)
Syntactic vectors are best for parsing (Bansal et al., 2014; Levy and Goldberg, 2014)

Don’t need huge unlabeled corpora for these methods to be effective
Results: English Treebank (Test)

Test set F1 all lengths

Neural+ Sparse: 91.1
Results: English Treebank (Test)

Test set F₁ all lengths

- **Neural+ Sparse**: 91.1
- **Sparse**: 89.2
Results: English Treebank (Test)

Test set F₁ all lengths

- Neural+ Sparse: 91.1
- Sparse: 89.2
- Berkeley Petrov+ 06: 90.1
Results: English Treebank (Test)

Test set $F_1$ all lengths

- **Neural+ Sparse**: 91.1
- **CCK Carreras+ 08**: 91.1
- **Sparse**: 89.2
- **Berkeley Petrov+ 06**: 90.1
Results: English Treebank (Test)

Test set F$_1$ all lengths

- Neural+: 91.1
- Sparse: 89.2
- Berkeley (Petrov+ 06): 90.1
- CCK (Carreras+ 08): 91.1
- ZPar (Zhu+ 13): 91.3
Results: English Treebank (Test)

Test set $F_1$ all lengths

- Neural+ Sparse: 91.1
- Sparse: 89.2
- Berkeley: 90.1
- CCK: 91.1
- ZPar: 91.3

Methods:
- Neural+ Sparse
- Sparse
- Berkeley
- CCK: Carreras+ 08
- ZPar: Zhu+ 13
Results: English Treebank (Test)

Test set $F_1$ all lengths

<table>
<thead>
<tr>
<th>Method</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Neural+ Sparse</td>
<td>91.1</td>
</tr>
<tr>
<td>Sparse</td>
<td>89.2</td>
</tr>
<tr>
<td>Berkeley</td>
<td>90.1</td>
</tr>
<tr>
<td>Petrov+ 06</td>
<td></td>
</tr>
<tr>
<td>CCK Carreras+ 08</td>
<td>91.1</td>
</tr>
<tr>
<td>ZPar Zhu+ 13</td>
<td>91.3</td>
</tr>
<tr>
<td>CVG Socher+ 13</td>
<td>90.4</td>
</tr>
</tbody>
</table>

(reranking ensemble)
Related Work

- Transition-based neural parsers: Henderson (2003), Chen and Manning (2014)
Related Work

- Transition-based neural parsers: Henderson (2003), Chen and Manning (2014)
- Local decisions only: Belinkov et al. (2014)
Related Work

- Transition-based neural parsers: Henderson (2003), Chen and Manning (2014)
- Local decisions only: Belinkov et al. (2014)
- Sequence-to-sequence LSTM: Vinyals et al. (2014)
Results: Other Languages

- Nine morphologically-rich languages from the SPMRL shared task
Results: Other Languages

- Nine morphologically-rich languages from the SPMRL shared task
- Word vectors trained on SPMRL monolingual data with word2vec (approximately 100M tokens per language)
Results: Other Languages

- Hall et al. (2014): 83.2
- Sparse+Neural: 85.1

Average F₁ scores on test set for all languages.
Results: Other Languages

- Hall et al. (2014)
- Sparse+Neural

Test set $F_1$ all lengths

Average: 83.2 vs. 85.1
Results: Other Languages

Test set F$_1$ all lengths

<table>
<thead>
<tr>
<th>Language</th>
<th>Hall et al. (2014)</th>
<th>Sparse+Neural</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average</td>
<td>83.2</td>
<td>85.1</td>
</tr>
<tr>
<td>Arabic</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Basque</td>
<td></td>
<td></td>
</tr>
<tr>
<td>French</td>
<td></td>
<td></td>
</tr>
<tr>
<td>German</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hebrew</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hungarian</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Korean</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Polish</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Swedish</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Hall et al. (2014) and Sparse+Neural performance comparison for test set F$_1$ across different languages.
Results: Other Languages

- Arabic
- Basque
- French
- German
- Hebrew
- Hungarian
- Korean
- Polish
- Swedish

Test set $F_1$ all lengths:

- Hall et al. (2014)
- Sparse+Neural

- Average $F_1$: 85.1
- Arabic $F_1$: 83.2
- Basque $F_1$: 85.1
- French $F_1$: 83.2
- German $F_1$: 85.1
- Hebrew $F_1$: 85.1
- Hungarian $F_1$: 85.1
- Korean $F_1$: 85.1
- Polish $F_1$: 85.1
- Swedish $F_1$: 85.1

- Works well even on smaller treebanks
Conclusion
Conclusion

- Neural nets can combine with CRFs to provide continuous features in discrete structured models
Neural nets can combine with CRFs to provide continuous features in discrete structured models

\[ S \]
Neural nets can combine with CRFs to provide continuous features in discrete structured models.

He gave a long speech on foreign policy.
Neural nets can combine with CRFs to provide continuous features in discrete structured models.

He gave a long speech on foreign policy.
Neural nets can combine with CRFs to provide continuous features in discrete structured models.
Neural nets can combine with CRFs to provide continuous features in discrete structured models.

He gave a long speech on foreign policy.
Neural nets can combine with CRFs to provide continuous features in discrete structured models
Neural nets can combine with CRFs to provide continuous features in discrete structured models.

Inference and learning are unchanged from the purely discrete model.
Conclusion

- Neural nets can combine with CRFs to provide continuous features in discrete structured models
- Inference and learning are unchanged from the purely discrete model
- High performance on constituency parsing for a range of languages
Conclusion

- Neural nets can combine with CRFs to provide continuous features in discrete structured models.

- Inference and learning are unchanged from the purely discrete model.

- High performance on constituency parsing for a range of languages.

nlp.cs.berkeley.edu/projects/neuralcrf.shtml
Conclusion

- Neural nets can combine with CRFs to provide continuous features in discrete structured models.
- Inference and learning are unchanged from the purely discrete model.
- High performance on constituency parsing for a range of languages.

[nlp.cs.berkeley.edu/projects/neuralcrf.shtml](nlp.cs.berkeley.edu/projects/neuralcrf.shtml)

Thank you!