Simple, Accurate Parsing with an All-Fragments Grammar

Mohit Bansal and Dan Klein
UC Berkeley
Independence Assumptions of PCFG

Strong Independence - lexical selection or agreement?

Fragments

Bod (1993)
Goodman (1996)

Refinement

Collins (1999)
Johnson (1998)
Petrov et al. (2006)

This Work (SIMPLE!)
All-Fragments Grammar $G$

**DERIVATIONS**

$X$

A
B
C

$\omega(d) = \prod_{f \in d} \omega(f)$

$\
\omega(d) = \prod_{f \in d} \omega(f)$

$\
t_{\text{max}} = \arg\max_t \sum_{d \in t} \omega(d)$

**FRAGMENTS**

$X$

Y
Z

A
B
C

exponential # of rules!!

(words)}
Fragment (data-oriented) Approach

The cat ate food

# of fragments = exponential in length of sentence!!

<table>
<thead>
<tr>
<th>Depth 1, 2 Subtrees</th>
<th># of Nodes</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.70</td>
<td>0.48</td>
</tr>
</tbody>
</table>

The diagram shows a tree structure with nodes labeled as follows:

- S (Sentence)
- NP (Noun Phrase)
  - DT (Determiner) - The
  - NN (Noun) - cat
- VP (Verb Phrase)
  - VB (Verb) - ate
  - NN (Noun) - food

The number of fragments is shown to be exponential in the length of the sentence.
$G^I$ - Implicit Representation of $G$

SYMBOLS:
- Base: $X$
- Indexed: $X_i$

RULES:
- Continue: $X_i \rightarrow Y_j \ Z_k$
- End: $X_i \rightarrow X$
- Begin: $X \rightarrow X_i$

Goodman (1996)

# of rules = |treebank B|
### Example

#### Training Data

<table>
<thead>
<tr>
<th>NP-1</th>
<th>DT-2</th>
<th>NN-3</th>
</tr>
</thead>
<tbody>
<tr>
<td>The</td>
<td></td>
<td></td>
</tr>
<tr>
<td>NP-4</td>
<td>DT-5</td>
<td>NN-6</td>
</tr>
<tr>
<td>A</td>
<td></td>
<td>dog</td>
</tr>
</tbody>
</table>

#### Rules

<table>
<thead>
<tr>
<th>NP-1</th>
<th>DT-2</th>
<th>NN-3</th>
</tr>
</thead>
<tbody>
<tr>
<td>The</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>NP-4</th>
<th>DT-5</th>
<th>NN-6</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td></td>
<td>dog</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>DT-2</th>
<th>NN-3</th>
</tr>
</thead>
<tbody>
<tr>
<td>DT</td>
<td>NN</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>DT-5</th>
<th>NN-6</th>
</tr>
</thead>
<tbody>
<tr>
<td>DT</td>
<td>NN</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>DT-2</th>
<th>NN-3</th>
</tr>
</thead>
<tbody>
<tr>
<td>DT</td>
<td>NN</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>CONT</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>END</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>BEGIN</th>
</tr>
</thead>
</table>
Parsing a Novel Test Sentence

Training Data

NP-1
  DT-2  NN-3
  The  cat

NP-4
  DT-5  NN-6
  A  dog

Test Parse

NP
  DT
  The

  NN
  dog

NP-1
  CONT

DT-2
  CONT

  NN-3
  BEGIN

  NN

  NN-6
  END

Derivation 1

NP-4
  CONT

DT-5
  BEGIN

  DT-2
  CONT

  The

  dog

Derivation 2

Novel Test Sentence

The  dog
Equivalence of $G$ and $G^I$

- Each derivation $d$ in $G$ reproducible in $G^I$

- Multiple derivations in $G^I$ correspond to same $d$ in $G$, differing in indices
Weights for Implicit Grammar

RULES

► CONT: \( X_i \rightarrow Y_j \ Z_k \)

► END: \( X_i \rightarrow X \)

► BEGIN: \( X \rightarrow X_i \)

WEIGHTS

\( \omega_{BODY} \ (\omega_{LEX}) \)

\( \omega_{SWITCH} \)

\[ \frac{1}{\# \text{ frags rooted at } X} \]

JUST 3 PARAMETERS!
words split into characters
Coarse-to-Fine Inference

coarse: ...

QP NP VP ...

refined: QP1 QP2 QP3

\[
\frac{P_{IN}(X, i, j) \cdot P_{OUT}(X, i, j)}{P_{IN}(\text{root}, 0, n)} < \text{threshold}
\]

Charniak et al. (2005, 2006)
Coarse-to-Fine Inference

"Fine" Grammar

NP-1
  DT-2  NN-3
NP-4
  DT-5  NN-6

"Coarse" Grammar

NP
  DT  NN

For same accuracy,
- 40x speed up
- 10x memory reduction

AVERAGE OVER INDICES

PCFG

The
Packed Graph Encoding

Tree-to-graph encoding

S

NP

The model

VBD

parsed

NP

a sentence

VP

NP

a sentence

VBD

was

PP

in the treebank

S

NP

The model

VBD

parsed

VP

NP

a sentence

VBD

was

PP

in the treebank
Savings from Packed Graph Encoding

- 1.4x speed up
- memory-usage < 4GB

- 20x speed up
- memory-usage < 8GB

Word-level Parsing

Indexed symbols (million)

Trees: 1.90
Graph: 0.90

Character-level Parsing

Indexed symbols (million)

Trees: 12.28
Graph: 1.11
Basic Refinement

The cat

NP

DT ADJ NN

P = 1

The small cat

NP

DT^NP ADJ^NP NN^NP

H = 1

The small cat

NP

DT^NP ADJ^NP NN^NP

@NP→_DT^NP

small

cat
Fragments Complement Refinements

<table>
<thead>
<tr>
<th>Model Type</th>
<th>F1 (dev ≤ 40)</th>
</tr>
</thead>
<tbody>
<tr>
<td>No-Refine (Raw PCFG)</td>
<td>71.3</td>
</tr>
<tr>
<td>Basic-Refine (P=H=1)</td>
<td>80.0</td>
</tr>
<tr>
<td>All-Frag + No-Refine</td>
<td>85.7</td>
</tr>
<tr>
<td>All-Frag + Basic-Refine</td>
<td>88.4</td>
</tr>
</tbody>
</table>
## Parsing Accuracy

### Word-level Parsing F1

<table>
<thead>
<tr>
<th>Decoding Objective</th>
<th>dev (≤ 40)</th>
<th>test (≤ 40)</th>
<th>test (all)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Max- Constituent$^1$</td>
<td>88.4</td>
<td>88.5</td>
<td>87.6</td>
</tr>
</tbody>
</table>

### Character-level Parsing F1

<table>
<thead>
<tr>
<th>Decoding Objective</th>
<th>dev (≤ 40)</th>
<th>test (≤ 40)</th>
<th>test (all)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Max-Constituent$^1$</td>
<td>88.2</td>
<td>88.0</td>
<td>87.1</td>
</tr>
</tbody>
</table>

$^1$Goodman (1996)
Full-scale Parsing

*word-level parsing results on dev-set (≤ 40)
Final WSJ Results

<table>
<thead>
<tr>
<th>Method</th>
<th>F1 (test ≤ 40)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Post and Gildea (2009)</td>
<td>82.6</td>
</tr>
<tr>
<td>Zuidema (2007)</td>
<td>83.8</td>
</tr>
<tr>
<td>Cohn et al. (2009)</td>
<td>84.0</td>
</tr>
<tr>
<td>All-Frag + Basic Refine</td>
<td>88.5</td>
</tr>
<tr>
<td>All-Frag + Addn Refine</td>
<td>88.7</td>
</tr>
</tbody>
</table>

*word-level parsing
*Addn Refine = Deterministic (NON-HEAD) annotation of Klein and Manning (2003)
*Cohn et al. - test all Zuidema - dev ≤ 100
Final WSJ Results

<table>
<thead>
<tr>
<th></th>
<th>F1 (test ≤ 40)</th>
</tr>
</thead>
<tbody>
<tr>
<td>All-Frag + Basic Refine</td>
<td>88.5</td>
</tr>
<tr>
<td>All-Frag + Addn Refine</td>
<td>88.7</td>
</tr>
<tr>
<td>Collins (1999)</td>
<td>88.6</td>
</tr>
<tr>
<td>Petrov and Klein (2007)</td>
<td>90.6</td>
</tr>
</tbody>
</table>

- This Paper
- Refinement-based Parsers
Other Language Results

**German**
- This Work: 79.8
- Dubey (2005): 76.3
- Petrov and Klein (2008): 81.5

**French**
- This Work: 78.0
- Arun and Keller (2005): 78.9
- Petrov and Klein (2008): 80.1

*F1 (test ≤ 40)*
Conclusions

- Practical, full-scale parsing with an all fragments grammar
  - Indexed grammar boils down to only 2-3 hyperparameters
  - Practical with natural coarse-to-fine projections and graph encodings
- Fragments complement refinements
  - Simple refinement + fragments $F1 \approx$ Collins 99
  - Accurate without an explicit lexicon
  - Zero training

<table>
<thead>
<tr>
<th>Parsing Model</th>
<th>F1 (test ≤ 40)</th>
<th>F1 (test all)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Collins (1999)</td>
<td>88.6</td>
<td>88.2</td>
</tr>
<tr>
<td>Our Model</td>
<td>88.7</td>
<td>88.1</td>
</tr>
</tbody>
</table>
Thank you!

Questions?

Berkeley

N L P