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?

This Work (SIMPLE!)

Bod (1993)
Goodman (1996)

Collins (1999)
Johnson (1998)
Petrov et al. (2006)
All-Fragments Grammar $G$

DERIVATIONS

FRAGMENTS

exponential # of rules!!

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

$t_{max} = \arg\max_t \sum_{d \in t} \omega(d)$
Fragment (data-oriented) Approach

The number of fragments is 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 = $|\text{treebank B}|$
Example

Training Data

Rules

NP-1

DT-2

NN-3

The

cat

NP-4

DT-5

NN-6

A

dog

NP-1

DT-2

NN-3

The

cat

NP-4

DT-5

NN-6

A

dog

DT-2

NN-3

DT-5

NN-6

DT

NN

END

DT-2

NN-3

DT-5

NN-6

DT

NN

BEGIN

CONT
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  NN
       The  dog

NP-1
   DT-2   NN-3
       The  dog

NP-4
   DT-5   NN-6
       The  dog

Novel Test Sentence

The dog

Derivation 1

Derivation 2
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!
Character-level Parsing

words split into characters
Coarse-to-Fine Inference

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

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

“Fine” Grammar

```
NP-1
  \(\text{DT-2} \quad \text{NN-3}\)
  \(\text{NP-4}\)
    \(\text{DT-5} \quad \text{NN-6}\)
```

“Coarse” Grammar

```
NP
  \(\text{DT} \quad \text{NN}\)
```

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

AVERAGE OVER INDICES

PCFG
Packed Graph Encoding

Tree-to-graph encoding
Savings from Packed Graph Encoding

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

Word-level Parsing

Indexed symbols (million)

Trees  | Graph
1.90   | 0.90

Character-level Parsing

Indexed symbols (million)

Trees  | Graph
12.28  | 1.11

− 20x speed up
− memory-usage < 8GB
Fragments Complement Refinements

| Condition                          | F1 (dev ≤ 40) |
|-----------------------------------|---------------|
| No-Refine (Raw PCFG)              | 71.3          |
| Basic-Refine (P=H=1)              | 80.0          |
| All-Frag + No-Refine              | 85.7          |
| All-Frag + Basic-Refine           | 88.4          |
### Parsing Accuracy

| Decoding Objective   | dev (≤ 40) | test (≤ 40) | test (all) |
|----------------------|------------|-------------|------------|
| Max-Constituent¹     | **88.4**   | **88.5**    | **87.6**   |

| Decoding Objective   | dev (≤ 40) | test (≤ 40) | test (all) |
|----------------------|------------|-------------|------------|
| Max-Constituent¹     | **88.2**   | **88.0**    | **87.1**   |

¹Goodman (1996)
Full-scale Parsing

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

F1 (test ≤ 40)

- Post and Gildea (2009): 82.6
- Zuidema (2007): 83.8
- Cohn et al. (2009): 84.0
- All-Frag + Basic Refine: 88.5
- All-Frag + Addn Refine: 88.7

*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

| Method                        | F1 Score |
|-------------------------------|----------|
| All-Frag + Basic Refine       | 88.5     |
| All-Frag + Addn Refine        | 88.7     |
| Collins (1999)                | 88.6     |
| Petrov and Klein (2007)       | 90.6     |

- **This Paper**
- **Refinement-based Parsers**
Other Language Results

German

|                | This Work | Dubey (2005) | Petrov and Klein (2008) |
|----------------|-----------|---------------|-------------------------|
| F1 (test ≤ 40) | 79.8      | 76.3          | 81.5                    |

French

|                | This Work | Arun and Keller (2005) | Petrov and Klein (2008) |
|----------------|-----------|------------------------|-------------------------|
| F1 (test ≤ 40) | 78.0      | 78.9                   | 80.1                    |
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 ≈ Collins 99
  - Accurate without an explicit lexicon
  - Zero training

| Parsing Model   | F1 (test ≤ 40) | F1 (test all) |
|-----------------|----------------|---------------|
| Collins (1999)  | 88.6           | 88.2          |
| Our Model       | 88.7           | 88.1          |
Thank you!

Questions?

Berkeley

NLP