Alignment and Marginal Inference with SCFGs

April 12, 2014
Forced decoding

• Other names
  • Forced alignment
  • Alignment
  • Synchronous parsing
  • Three-way composition

• Why?
  • Compute expectations of
\begin{align*}
\text{Items} & \quad [X, s, t, u, v] \\
\text{Axioms} & \quad [X, s, s + 1, u, u + 1] \quad X \rightarrow \langle f_s, e_u \rangle \in G \\
& \quad [X, s, s + 1, u, u] \quad X \rightarrow \langle f_s, \varepsilon \rangle \in G \\
& \quad [X, s, s, u, u + 1] \quad X \rightarrow \langle \varepsilon, e_u \rangle \in G
\end{align*}
**Wu, 1997**

**Rules**

\[
\frac{[X, s, j, u, k] \ [Y, j, t, k, v]}{[Z, s, t, u, v]} \quad Z \rightarrow \langle X \ Y, \bar{1} \bar{2} \rangle \in G
\]

\[
\frac{[X, s, j, k, v] \ [Y, j, t, u, k]}{[Z, s, t, u, v]} \quad Z \rightarrow \langle X \ Y, \bar{2} \bar{1} \rangle \in G
\]

**Goal**

\[ [S, 1, 1, m + 1, n + 1] \]
Grammar

NP → ⟨NN JJ, [2 1]⟩

NN → ⟨bruja, witch⟩

JJ → ⟨verde, green⟩

Items
Grammar

\[ NP \rightarrow \langle NN \ JJ, \begin{bmatrix} 2 \\ 1 \end{bmatrix} \rangle \]

\[ NN \rightarrow \langle bruja, \ witch \rangle \]

\[ JJ \rightarrow \langle verde, \ green \rangle \]

Items

\[ \{NN, \ 0, \ 1, \ 1, \ 2\} \]
Grammar

NP $\rightarrow \langle NN \ JJ, \begin{bmatrix} 2 & 1 \end{bmatrix} \rangle$

NN $\rightarrow \langle bruja, \ witch \rangle$

JJ $\rightarrow \langle verde, \ green \rangle$

Items

\[ [NN, 0, 1, 1, 2] \]

\[ [JJ, 1, 2, 0, 1] \]
**Grammar**

NP → ⟨NN JJ, 2 1⟩

NN → ⟨bruja, witch⟩

JJ → ⟨verde, green⟩

**Items**

([NN, 0, 1, 1, 2] [JJ, 1, 2, 0, 1]) → ([X, s, j, k, v] [Y, j, t, u, k]

[Z, s, t, u, v] Z → ⟨X Y, 2 1⟩ ∈ G)
Grammar

NP → ⟨NN JJ, [2 1]⟩
NN → ⟨bruja, witch⟩
JJ → ⟨verde, green⟩

Items

[NN, 0, 1, 1, 2]
[JJ, 1, 2, 0, 1]
[NP, 0, 2, 0, 2]
Wu, 1997 - Analysis

Rules

$$\frac{[X, s, j, u, k] \ [Y, j, t, k, v]}{[Z, s, t, u, v]}$$

$$Z \rightarrow \langle X \ Y, 1 \ 2 \rangle \in G$$

$$\frac{[X, s, j, k, v] \ [Y, j, t, u, k]}{[Z, s, t, u, v]}$$

$$Z \rightarrow \langle X \ Y, 2 \ 1 \rangle \in G$$

Run-time: $O(n^6)$
Can we do better?
Hypergraph review

la lectura : reading

ayer : yesterday

Source label

Target label

Goal node
Hypergraph review

la lectura : reading  
1 de 2 : 2's 1

ayer : yesterday  
1 de 2 : 1 from 2

Substitution sites / variables / non-terminals
Hypergraphs as Grammars

• **Claim:** A hypergraph is isomorphic to a (synchronous) CFG

• LM integration can be understood as the intersection of an regular and CF language

• Cube pruning approximates this intersection
Two-Parse Algorithm
(Dyer, 2010)
Input: <dianzi shiang de mao, a cat on the mat>

With thanks and apologies to Zhifei Li.
Input: <dianzi shiang de mao, a cat on the mat>

With thanks and apologies to Zhifei Li.
Input: <"dianzi shiang de mao , a cat on the mat">
Input: <dianzi shiang de mao , a cat on the mat>

Isomorphic CFG

[X34] → a cat
[X02] → the mat
Input: <dianzi shiang de mao , a cat on the mat>

Isomorphic CFG

[X34] → a cat
[X02] → the mat
[X04a] → [X34] on [X02]
[X04a] → [X34] of [X02]
Input: <dianzi shiang de mao, a cat on the mat>

Isomorphic CFG

[X34] → a cat
[X02] → the mat
[X04a] → [X34] on [X02]
[X04a] → [X34] of [X02]
[X04b] → [X02] 's [X34]
[X04b] → [X02] [X34]
Input: <dianzi shiang de mao, a cat on the mat>

Isomorphic CFG

[X34] → a cat
[X02] → the mat
[X04a] → [X34] on [X02]
[X04a] → [X34] of [X02]
[X04b] → [X02] ‘s [X34]
[X04b] → [X02] [X34]
[S] → [X04a]
[S] → [X04b]
Input: <dianzi shiang de mao, a cat on the mat>

**Isomorphic CFG**

\[X34] \rightarrow a\ cat
\[X02] \rightarrow the\ mat
\[X04a] \rightarrow [X34] \text{on} [X02]
\[X04a] \rightarrow [X34] \text{of} [X02]
\[X04b] \rightarrow [X02] \text{‘s} [X34]
\[X04b] \rightarrow [X02] [X34]
[S] \rightarrow [X04a]
[S] \rightarrow [X04b]
Isomorphic CFG

\[
\begin{align*}
[X34] & \rightarrow a \text{ cat} \\
[X02] & \rightarrow the \text{ mat} \\
[X04a] & \rightarrow [X34] \text{ on } [X02] \\
[X04a] & \rightarrow [X34] \text{ of } [X02] \\
[X04b] & \rightarrow [X02] \text{'s } [X34] \\
[X04b] & \rightarrow [X02] [X34] \\
[S] & \rightarrow [X04a] \\
[S] & \rightarrow [X04b]
\end{align*}
\]

\text{ Input: } <\textit{dianzi shiang de mao, a cat on the mat}>
Input: <dianzi shiang de mao, a cat on the mat>

**Isomorphic CFG**

\[
\begin{align*}
[X34] & \rightarrow \text{a cat} \\
[X02] & \rightarrow \text{the mat} \\
[X04a] & \rightarrow [X34] \text{ on } [X02] \\
[X04a] & \rightarrow [X34] \text{ of } [X02] \\
[X04b] & \rightarrow [X02] \text{'s } [X34] \\
[X04b] & \rightarrow [X02] [X34] \\
[S] & \rightarrow [X04a] \\
[S] & \rightarrow [X04b]
\end{align*}
\]

\[
[X34] \\
\downarrow \\
a \text{ cat} \quad \text{on} \quad \text{the mat}
\]
Input: <dianzi shiang de mao, a cat on the mat>

**Isomorphic CFG**

\[
\begin{align*}
[X34] & \to \text{a cat} \\
[X02] & \to \text{the mat} \\
[X04a] & \to [X34] \text{ on } [X02] \\
[X04a] & \to [X34] \text{ of } [X02] \\
[X04b] & \to [X02] \text{'s } [X34] \\
[X04b] & \to [X02] [X34] \\
[S] & \to [X04a] \\
[S] & \to [X04b]
\end{align*}
\]
Input: <dianzi shiang de mao, a cat on the mat>

**Isomorphic CFG**

- [X34] → a cat
- [X02] → the mat
- [X04a] → [X34] on [X02]
- [X04a] → [X34] of [X02]
- [X04b] → [X02] ‘s [X34]
- [X04b] → [X02] [X34]
- [S] → [X04a]
- [S] → [X04b]
Input: <dianzi shiang de mao , a cat on the mat>

Isomorphic CFG

[X34] → a cat
[X02] → the mat
[X04a] → [X34] on [X02]
[X04a] → [X34] of [X02]
[X04b] → [X02] ‘s [X34]
[X04b] → [X02] [X34]
[S] → [X04a]
[S] → [X04b]
Two Algorithms

- **ITG algorithm** *(Wu, 1997)*
  - Jointly parse both source and target
  - Only works for binary ITGs (although generalizable)
  - Runs in \( \Theta(n^6) \)

- **Two-parse algorithm** *(Dyer, 2010)*
  - Parse source, then parse target
  - Works with any SCFG
  - For binary ITGs runs in \( O(n^6) \)
In the first experiment, we compare performance of the two-parse algorithm and the ITG parsing algorithm on an Arabic-English phrasal ITG alignment task. The corpus consisted of 0.9k sentence pairs, involving 96M Arabic tokens and 114M English tokens, drawn from the NIST MT evaluation newswire training data. Sentences were filtered to a length of maximally 12 tokens on either side. For this experiment, we used a variant of the phrasal ITG described by Zhang et al. The restriction that phrases contain exactly a single alignment point was relaxed; instead, the grammar was restricted to contain all phrases consistent with the word-based alignment up to a maximal phrase size of 6. This resulted in a synchronous grammar with 8.69M rules. Figure 60 plots the average runtime of the two algorithms as a function of the Arabic sentence length. Table 6- shows the overall average runtimes. Both presentations make clear that the two-parse approach is dramatically more efficient. In total, aligning the 0.9k sentence pairs in the corpus completed in less than 0.5 hours with the two-parse algorithm but required more than 0.1 week with the baseline ITG parsing algorithm.

Table 6-: Comparison of synchronous parsing algorithms on Arabic-English.

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>avg runtime (sec)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ITG alignment</td>
<td>-0.9</td>
</tr>
<tr>
<td>Two-parse algorithm</td>
<td>9.69</td>
</tr>
</tbody>
</table>

In the second experiment, we evaluate an alternative approach to computing a synchronous parse forest that is based on cube pruning by Huang and Chiang. A note on implementation: our ITG aligner was minimal; it only computed the probability of the sentence pair using the inside algorithm. With the two-parse aligner, we stored the complete item chart during both the first and second parses. Therefore the implementation was biased in favor of the baseline ITG parsing algorithm.