Questions for PCFGs

3 questions for Probabilistic Context Free Grammars (PCFGs):

- What is the probability of a sentence $w_{1m}$ according to grammar $G$? $P(w_{1m} \mid G)$
- What is the most likely parse for a sentence? $\arg\max_{P(t \mid w_{1m}, G)}$
- How can we choose rule probabilities for the grammar $G$ that maximize the probability of a sentence? $\arg\max_G P(w_{1m} \mid G)$

Calculating $P(w_{1m})$

take one subtree from $w_p$ to $w_q$:

- **outside probability**: probability of beginning with start symbol $N^i$ and generating the nonterminal $N_{pq}^j$ and all the words not in the subtree, $a_j(p, q) = P(w_{1(p-1)}, N_{pq}^j, w_{(q+1)m} \mid G)$
- **inside probability**: probability of words $w_p \ldots w_q$ given that $N^i$ is the starting point, $\beta_j(p, q) = P(w_{pq} \mid N_{pq}^j, G)$
Calculating $P(w_1m)$ (2)

Inside Probabilities – Base Case

- **Base case:** subtree for word $w_k$ and rule $N_j \rightarrow w_k$

  $$\beta_j(k, k) = P(w_k | N_j^k, G) = P(N_j \rightarrow w_k | G)$$

Inside Probabilities – Induction

- **Induction:** find $\beta_j(p, q)$ for $p < q$

  Chomsky Normal Form ⇒ rule of form: $N_j \rightarrow N^l \ N^s$

  $$\beta_j(p, q) = \sum_{r,s} \sum_{d=p}^{q-1} P(N_j \rightarrow N^l \ N^s) \times \beta_r(p, d) \times \beta_s(d + 1, q)$$

The inside algorithm

- When at a particular focus non-terminal node $N_j$, figure out the different ways to expand the node.
- Implementation: only consider expansions which match a previously calculated inside probability

So, for example, in a sentence with a VP trying to cover positions 2 through 5 (astronomers [saw stars with ears]):

$\beta_{VP}(2, 5) = P(VP \rightarrow V \ NP)\beta_{V}(2, 2)\beta_{NP}(3, 5) + P(VP \rightarrow VP \ PP)\beta_{VP}(2, 3)\beta_{PP}(4, 5)$

Probability of a string

$$P(w_1m | G) = P(N_1 \Rightarrow^* w_1m | G) = P(w_1m | N_1^m, G) = \beta_1(1, m)$$
**Outside Probabilities – Induction**

**base case:** probability of root being $N_1^1$, nothing outside

\[
\alpha_1(1, m) = 1 \quad N_1^1 \text{ is the start symbol}
\]

\[
\alpha_j(1, m) = 0 \quad \text{for } j \neq 1
\]

**Outside Probabilities – Induction (2)**

**induction:** node $N_{pq}^l$ can be either the left or the right daughter \(\Rightarrow\) sum over both possibilities

\[
\alpha_j(p, q) = \sum_{l \in \{\text{left}, \text{right}\}} \sum_{e=q+1}^{m} P(w_l(p-1), w_l(q+1)m \mid N_{pe}^l, N_{pq}^l, N_{e(q+1)e}^l)
\]

\[
+ \sum_{l \in \{\text{left}, \text{right}\}} \sum_{e=q+1}^{m} P(w_{e(e-1)}, w_l(q+1)m \mid N_{eq}^l, N_{pq}^l, N_{e(q+1)e}^l)
\]

\[
= \sum_{l \in \{\text{left}, \text{right}\}} \sum_{e=q+1}^{m} P(w_l(p-1), w_l(q+1)m \mid N_{pe}^l)
\]

\[
\times P(N_{pq}^l, N_{e(q+1)e}^l \mid N_{pe}^l, P(w_l(q+1)m \mid N_{eq}^l, N_{pq}^l, N_{e(q+1)e}^l)
\]

**Outside Probabilities – Induction (3)**

**induction:** node $N_{pq}^l$ can be either the left or the right daughter \(\Rightarrow\) sum over both possibilities

\[
\alpha_j(p, q) = \sum_{l \in \{\text{left}, \text{right}\}} \sum_{e=q+1}^{m} P(w_l(p-1), w_l(q+1)m \mid N_{pe}^l)
\]

\[
\times P(N_{pq}^l, N_{e(q+1)e}^l \mid N_{pe}^l, P(w_l(q+1)m \mid N_{eq}^l, N_{pq}^l, N_{e(q+1)e}^l)
\]

**Outside Probabilities – Induction (4)**

**induction:** node $N_{pq}^l$ can be either the left or the right daughter \(\Rightarrow\) sum over both possibilities

\[
\alpha_j(p, q) = \sum_{l \in \{\text{left}, \text{right}\}} \sum_{e=q+1}^{m} P(w_l(p-1), w_l(q+1)m \mid N_{pe}^l)
\]

\[
\times P(N_{pq}^l, N_{e(q+1)e}^l \mid N_{pe}^l, P(w_l(q+1)m \mid N_{eq}^l, N_{pq}^l, N_{e(q+1)e}^l)
\]

+ \sum_{l \in \{\text{left}, \text{right}\}} \sum_{e=q+1}^{m} P(w_{e(e-1)}, w_l(q+1)m \mid N_{eq}^l, N_{pq}^l, N_{e(q+1)e}^l)

\[
\times P(N_{pq}^l, N_{e(q+1)e}^l \mid N_{eq}^l, P(w_{e(e-1)}, w_l(q+1)m \mid N_{eq}^l, N_{pq}^l, N_{e(q+1)e}^l)
\]

\[
= \sum_{l \in \{\text{left}, \text{right}\}} \sum_{e=q+1}^{m} P(w_l(p-1), w_l(q+1)m \mid N_{pe}^l)
\]

\[
\times P(N_{pq}^l, N_{e(q+1)e}^l \mid N_{pe}^l, P(w_l(q+1)m \mid N_{eq}^l, N_{pq}^l, N_{e(q+1)e}^l)
\]

+ \sum_{l \in \{\text{left}, \text{right}\}} \sum_{e=q+1}^{m} P(w_{e(e-1)}, w_l(q+1)m \mid N_{eq}^l, N_{pq}^l, N_{e(q+1)e}^l)

\[
\times P(N_{pq}^l, N_{e(q+1)e}^l \mid N_{eq}^l, P(w_{e(e-1)}, w_l(q+1)m \mid N_{eq}^l, N_{pq}^l, N_{e(q+1)e}^l)
**Outside Probabilities – Example**

**Sentence:** astronomers saw stars with ears

\( \alpha_S(1,5) = 1.0 \)

**Complete Table:**

<table>
<thead>
<tr>
<th>( k )</th>
<th>( \alpha_S(1,5) )</th>
<th>( \alpha_{NP}(2,5) )</th>
<th>( \alpha_{VP}(3,5) )</th>
<th>( \alpha_{PP}(4,5) )</th>
<th>( \alpha_{NP}(5,5) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>1.0</td>
<td>0.1</td>
<td>0.07</td>
<td>0.00882</td>
<td>0.00882</td>
</tr>
<tr>
<td>4</td>
<td>( \alpha_{PP}(4,4) = 0.015876 )</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>( \alpha_{VP}(2.3) = 0.0054 )</td>
<td>( \alpha_{NP}(3,3) = 0.00882 )</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>( \alpha_{V}(2.2) = 0.0015876 )</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>( \alpha_{NP}(1,1) = 0.015876 )</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**End Sentence:** astronomers saw stars with ears

**Probability of a string**

For any \( k, 1 \leq k \leq m \)

\[ P(w_{1:m} \mid G) = \sum_j P(w_{1:(k-1)}, w_k, w_{(k+1):m}, N_{kk}^j \mid G) \]

**Outside Probabilities – Example**

**Sentence:** astronomers saw stars with ears

\( \alpha_S(1,5) = 1.0 \)

**Complete Table:**

<table>
<thead>
<tr>
<th>( k )</th>
<th>( \alpha_S(1,5) )</th>
<th>( \alpha_{NP}(2,5) )</th>
<th>( \alpha_{VP}(3,5) )</th>
<th>( \alpha_{PP}(4,5) )</th>
<th>( \alpha_{NP}(5,5) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>1.0</td>
<td>0.1</td>
<td>0.07</td>
<td>0.00882</td>
<td>0.00882</td>
</tr>
<tr>
<td>4</td>
<td>( \alpha_{PP}(4,4) = 0.015876 )</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>( \alpha_{VP}(2.3) = 0.0054 )</td>
<td>( \alpha_{NP}(3,3) = 0.00882 )</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>( \alpha_{V}(2.2) = 0.0015876 )</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>( \alpha_{NP}(1,1) = 0.015876 )</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**End Sentence:** astronomers saw stars with ears

**Probability of a string**

For any \( k, 1 \leq k \leq m \)

\[ P(w_{1:m} \mid G) = \sum_j P(w_{1:(k-1)}, w_k, w_{(k+1):m}, N_{kk}^j \mid G) \]

\[ = \sum_j P(w_{1:(k-1)}, N_{kk}^j, w_{(k+1):m} \mid G) \times P(w_k \mid w_{(k-1)}, N_{kk}^j, w_{(k+1):m} \mid G) \]
**Probability of a string**

For any $1 \leq k \leq m$

$$P(w_{1m} \mid G) = \sum_j P(w_{1(k-1)}, w_k, w_{(k+1)}^m \mid G)$$

$$= \sum_j P(w_{1(k-1)}, N^j_{kk}, w_{(k+1)}^m \mid G) \times P(w_k \mid N^j_{kk}, w_{(k+1)}^m, G)$$

$$= \sum_j \alpha_j(k, k) P(N^i \rightarrow w_k)$$

**Finding the Most Likely Parse**

**Goal:** finding the parse with the highest probability:

$$P(\hat{t}) = \max_{i \leq j \leq m} \delta(i, m)$$

- **Method:** Viterbi-style parsing
  - Viterbi: defining accumulators $\delta_j(p, q)$ which record the highest probability for the subtree with root $N^j$ dominating the words $w_{pq}$
  - **Idea:** if we assume that a node $N_{pq}^j$ is used in the derivation, then:
    - it needs to be the subtree with maximal probability,
    - all other subtrees for $w_{pq}$ having the root $N^j$ have a lower probability, and will result in an analysis with lower probability

**Viterbi Parsing**

- **Initialization:**
  $$\delta_1(p, p) = P(N^1 \rightarrow w_p)$$

- **Induction:**
  $$\delta_j(p, q) = \max_{i \leq j \leq m \leq p} \delta_j(p, r) \times \delta_k(r + 1, q)$$

- **Store backtrace:**
  $$\Psi_j(p, q) = \arg \max_{i \leq j \leq m \leq p} \delta_j(p, r) \times \delta_k(r + 1, q)$$

**Viterbi Parsing (2)**

- **Termination:**
  $$P(\hat{t}) = \delta_1(1, m)$$

- **Reconstruction:**
  - root node: $N^1_{pq}$
  - general case: $N_{pq}^j \in \hat{t}$ and $\Psi_j(p, q) = (j, k, r)$
    - left($N_{pq}^j$) = $N_{pr}^j$
    - right($N_{pq}^j$) = $N_{(r+1)q}^j$
Probalistic CYK Parsing (2)

// base case – lexical rules
for i = 1 to n do
  for all $A \rightarrow w_i \in G$ do
    $\delta(i, 1, A) = P(A \rightarrow w_i)$

// recursive case
for len = 2 to n
  for i = 1 to n – len + 1
    for k = 1 to len – 1
      for all $A \rightarrow B C \in G$ do
        prob = $P(A \rightarrow B C) \times \delta(i, k, B) \times \delta(i + k, len - k, C)$
        if prob > $\delta(i, len, A)$ then
          $\delta(i, len, A) = prob$
          $\Psi(i, len, A) = (B, C, k)$

indexes of $\delta, \Psi$: (no. of first word of const., length of const., root node)
value of $\Psi$: (first daughter, second daughter, length of first daughter)
other: beginning of second daughter, length of second daughter

Probabilistic CYK Parsing – Example

Grammar:

<table>
<thead>
<tr>
<th>Non-terminal</th>
<th>Production</th>
<th>Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>S</td>
<td>PP MS</td>
<td>1.0</td>
</tr>
<tr>
<td></td>
<td>IN</td>
<td>0.5</td>
</tr>
<tr>
<td>PP</td>
<td>IN NP</td>
<td>1.0</td>
</tr>
<tr>
<td></td>
<td>in</td>
<td>0.5</td>
</tr>
<tr>
<td>MS</td>
<td>NP VP</td>
<td>0.8</td>
</tr>
<tr>
<td></td>
<td>JJ</td>
<td>1.0</td>
</tr>
<tr>
<td>MS</td>
<td>NN VP</td>
<td>0.2</td>
</tr>
<tr>
<td></td>
<td>NN</td>
<td>0.2</td>
</tr>
<tr>
<td>MS</td>
<td>NP VP</td>
<td>0.2</td>
</tr>
<tr>
<td></td>
<td>NN</td>
<td>0.3</td>
</tr>
<tr>
<td>NP</td>
<td>DT NN</td>
<td>0.5</td>
</tr>
<tr>
<td></td>
<td>NN</td>
<td>0.3</td>
</tr>
<tr>
<td>NP</td>
<td>JJ NN</td>
<td>0.3</td>
</tr>
<tr>
<td></td>
<td>NN</td>
<td>0.2</td>
</tr>
<tr>
<td>NP</td>
<td>NP PP</td>
<td>0.2</td>
</tr>
<tr>
<td></td>
<td>NN</td>
<td>0.3</td>
</tr>
<tr>
<td>NP</td>
<td>NP PP</td>
<td>0.7</td>
</tr>
<tr>
<td></td>
<td>VBD</td>
<td>1.0</td>
</tr>
<tr>
<td>VP</td>
<td>VBD NP</td>
<td>0.6</td>
</tr>
<tr>
<td></td>
<td>DT</td>
<td>0.6</td>
</tr>
<tr>
<td>VP</td>
<td>VBD XP</td>
<td>0.4</td>
</tr>
<tr>
<td></td>
<td>DT</td>
<td>0.4</td>
</tr>
<tr>
<td>XP</td>
<td>VBD XP</td>
<td>0.4</td>
</tr>
<tr>
<td></td>
<td>DT</td>
<td>0.4</td>
</tr>
<tr>
<td>XP</td>
<td>NN PP</td>
<td>0.3</td>
</tr>
</tbody>
</table>

Sentence:

At Roman banquets, the guests wore garlic in their hair