CS626-449: Speech, NLP and the Web/Topics in AI

Pushpak Bhattacharyya
CSE Dept.,
IIT Bombay
Lecture-17: Probabilistic parsing; inside-outside probabilities
Probability of a parse tree (cont.)

\[
P(t|s) = P(t | S_{1,l}) = P(\text{NP}_{1,2}, \text{DT}_{1,1}, w_1, \text{N}_{2,2}, w_2, \text{VP}_{3,l}, \text{V}_{3,3}, w_3, \text{PP}_{4,l}, \text{P}_{4,4}, w_4, \text{NP}_{5,l}, w_5\ldots | S_{1,l})
\]

\[
= P(\text{NP}_{1,2}, \text{VP}_{3,l} | S_{1,l}) \cdot P(\text{DT}_{1,1}, \text{N}_{2,2} | \text{NP}_{1,2}) \cdot D(w_1 | \text{DT}_{1,1}) \cdot P(w_2 | \text{N}_{2,2}) \cdot P(\text{V}_{3,3}, \text{PP}_{4,l} | \text{VP}_{3,l}) \cdot P(w_3 | \text{V}_{3,3}) \cdot P(\text{P}_{4,4}, \text{NP}_{5,l} | \text{PP}_{4,l}) \cdot P(w_4 | \text{P}_{4,4}) \cdot P(w_5\ldots | \text{NP}_{5,l})
\]

(Using Chain Rule, Context Freeness and Ancestor Freeness)

Example PCFG Rules & Probabilities

- $S \rightarrow NP \ VP \ 1.0$
- $NP \rightarrow DT \ NN \ 0.5$
- $NP \rightarrow NNS \ 0.3$
- $NP \rightarrow NP \ PP \ 0.2$
- $PP \rightarrow P \ NP \ 1.0$
- $VP \rightarrow VP \ PP \ 0.6$
- $VP \rightarrow VBD \ NP \ 0.4$

- $DT \rightarrow the \ 1.0$
- $NN \rightarrow gunman \ 0.5$
- $NN \rightarrow building \ 0.5$
- $VBD \rightarrow sprayed \ 1.0$
- $NNS \rightarrow bullets \ 1.0$
• The gunman sprayed the building with bullets.

\[
P(t_1) = 1.0 \times 0.5 \times 1.0 \times 0.5 \times 0.6 \times 0.4 \times 1.0 \times 0.5 \times 1.0 \times 0.5 \times 1.0 \\
= 0.00225
\]
Another Parse $t_2$

- The gunman sprayed the building with bullets.

\[
P(t_2) = 1.0 \times 0.5 \times 1.0 \times 0.5 \times 0.4 \times 1.0 \times 0.2 \times 0.5 \times 1.0 \times 0.5 \times 1.0 \times 0.3 \times 1.0 = 0.0015
\]
HMM ↔ PCFG

- O observed sequence ↔ $w_{1m}$ sentence
- X state sequence ↔ $t$ parse tree
- $\mu$ model ↔ G grammar

- Three fundamental questions
HMM ↔ PCFG

- How likely is a certain observation given the model? ↔ How likely is a sentence given the grammar?
  \[ P(O \mid \mu) \leftrightarrow P(w_{1m} \mid G) \]

- How to choose a state sequence which best explains the observations? ↔ How to choose a parse which best supports the sentence?
  \[ \arg \max_{X} P(X \mid O, \mu) \leftrightarrow \arg \max_{t} P(t \mid w_{1m}, G) \]
HMM ↔ PCFG

• How to choose the model parameters that best explain the observed data? ↔ How to choose rule probabilities which maximize the probabilities of the observed sentences?

\[
\arg\max_{\mu} P(O | \mu) \leftrightarrow \arg\max_{G} P(w_{1m} | G)
\]
Interesting Probabilities

What is the probability of having a NP at this position such that it will derive “the building” ? - $\beta_{NP}(4,5)$

Inside Probabilities

The gunman sprayed   the building   with bullets
1  2  3  4  5  6  7

What is the probability of starting from N$^1$ and deriving “The gunman sprayed”, a NP and “with bullets” ? - $\alpha_{NP}(4,5)$

Outside Probabilities
Outside Probabilities

- Forward ↔ Outside probabilities
- $\alpha_j(p,q) :$ The probability of beginning with $N_1^j$ & generating the non-terminal $N_{pq}^j$ and all words outside $w_p...w_q$
- Forward probability: $\alpha_i(t) = P(w_{1(t-1)}, X_t = i \mid \mu)$
- Outside probability: $\alpha_j(p,q) = P(w_{1(p-1)}, N_{pq}^j, w_{(q+1)m} \mid G)$
Inside Probabilities

- Backward ↔ Inside probabilities
- $\beta_j(p,q)$: The probability of generating the words $w_{p..q}$ starting with the non-terminal $N_{pq}^j$.

- Backward probability: $\beta_i(t) = P(w_{tT} \mid X_t = i, \mu)$
- Inside probability: $\beta_j(p,q) = P(w_{pq} \mid N_{pq}^j, G)$
Outside & Inside Probabilities

\[ \alpha_{NP}(4,5) \text{ for } "\text{the building}" \]

\[ = P(\text{The gunman sprayed, } NP_{4,5}, \text{ with bullets} \mid G) \]

\[ \beta_{NP}(4,5) \text{ for } "\text{the building}" = P(\text{the building} \mid NP_{4,5}, G) \]
Inside probabilities $\beta_j(p,q)$

Base case:

$$\beta_j(k,k) = P(w_k \mid N_{kk}^j, G) = P(N_{kk}^j \rightarrow w_k \mid G)$$

- Base case is used for rules which derive the words or terminals directly
  
  *E.g.*, Suppose $N^j = NN$ is being considered & $NN \rightarrow$ building is one of the rules with probability 0.5

  $$\beta_{NN}(5,5) = P(building \mid NN_{5,5}, G) = P(NN_{5,5} \rightarrow building \mid G) = 0.5$$
**Induction Step**

Consider different splits of the words - indicated by $d$

*E.g., the huge building*

Split here for $d=2$  
$d=3$

Consider different non-terminals to be used in the rule:

NP $\rightarrow$ DT NN, NP $\rightarrow$ DT NNS are available options

Consider summation over all these.

**Formula:**

$$\beta_j(p, q) = P(w_{pq} \mid N_{pq}^j, G)$$

$$= \sum_{r,s} \sum_{d=p}^{q-1} P(N^j \rightarrow N^r N^s) \cdot \beta_r(p, d) \cdot \beta_s(d+1, q)$$
The Bottom-Up Approach

• The idea of induction
• Consider “the gunman”

Base cases: Apply unary rules
- DT $\rightarrow$ the \hspace{2cm} Prob = 1.0
- NN $\rightarrow$ gunman \hspace{2cm} Prob = 0.5

Induction: Prob that a NP covers these 2 words
= P (NP $\rightarrow$ DT NN) * P (DT deriving the word “the”) * P (NN deriving the word “gunman”)
= 0.5 * 1.0 * 0.5 = 0.25
Parse Triangle

- A parse triangle is constructed for calculating $\beta_j(p,q)$
- Probability of a sentence using $\beta_j(p,q)$:

$$P(w_{1m} \mid G) = P(N^1 \rightarrow w_{1m} \mid G) = P(w_{1m} \mid N_{1m}^1, G) = \beta_1(1, m)$$
### Parse Triangle

<table>
<thead>
<tr>
<th></th>
<th>The (1)</th>
<th>gunman (2)</th>
<th>sprayed (3)</th>
<th>the (4)</th>
<th>building (5)</th>
<th>with (6)</th>
<th>bullets (7)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>$\beta_{DT} = 1.0$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td>$\beta_{NN} = 0.5$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
<td></td>
<td>$\beta_{VBD} = 1.0$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>$\beta_{DT} = 1.0$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>$\beta_{NN} = 0.5$</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>$\beta_{P} = 1.0$</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>$\beta_{NNS} = 1.0$</td>
</tr>
</tbody>
</table>

- Fill diagonals with $\beta_j(k, k)$
### Parse Triangle

<table>
<thead>
<tr>
<th></th>
<th>The (1)</th>
<th>gunman (2)</th>
<th>sprayed (3)</th>
<th>the (4)</th>
<th>building (5)</th>
<th>with (6)</th>
<th>bullets (7)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>$\beta_{DT} = 1.0$</td>
<td>$\beta_{NP} = 0.25$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td></td>
<td>$\beta_{NN} = 0.5$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>$\beta_{VBD} = 1.0$</td>
</tr>
<tr>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>$\beta_{DT} = 1.0$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>$\beta_{NN} = 0.5$</td>
</tr>
<tr>
<td>6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>$\beta_{P} = 1.0$</td>
</tr>
<tr>
<td>7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>$\beta_{NNS} = 1.0$</td>
</tr>
</tbody>
</table>

- **Calculate using induction formula**

$$\beta_{NP}(1, 2) = P(\text{the gunman} \mid NP_{1, 2}, G)$$

$$= P(NP \rightarrow DT \ NN) \cdot \beta_{DT}(1, 1) \cdot \beta_{NN}(2, 2)$$

$$= 0.5 \cdot 1.0 \cdot 0.5 = 0.25$$
The gunman sprayed the building with bullets.
Another Parse $t_2$

- The gunman sprayed the building with bullets.

Rule used here is

$VP \rightarrow VBD \ NP$
### Parse Triangle

<table>
<thead>
<tr>
<th></th>
<th>The (1)</th>
<th>gunman (2)</th>
<th>sprayed (3)</th>
<th>the (4)</th>
<th>building (5)</th>
<th>with (6)</th>
<th>bullets (7)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>$\beta_{DT} = 1.0$</td>
<td>$\beta_{NP} = 0.25$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>$\beta_s = 0.0465$</td>
</tr>
<tr>
<td>2</td>
<td></td>
<td></td>
<td>$\beta_{NN} = 0.5$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
<td></td>
<td>$\beta_{VBD} = 1.0$</td>
<td>$\beta_{VP} = 1.0$</td>
<td>$\beta_{VP} = 0.186$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td>$\beta_{DT} = 1.0$</td>
<td>$\beta_{NP} = 0.25$</td>
<td></td>
<td>$\beta_{NP} = 0.015$</td>
</tr>
<tr>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>$\beta_{NN} = 0.5$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>$\beta_p = 1.0$</td>
<td>$\beta_{PP} = 0.3$</td>
</tr>
<tr>
<td>7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>$\beta_{NNS} = 1.0$</td>
</tr>
</tbody>
</table>

$\beta_{VP}(3, 7) = P(\text{sprayed the building with bullets} \mid VP_{3,7}, G)$

$$= P(VP \rightarrow VP PP) \cdot \beta_{VP}(3, 5) \cdot \beta_{PP}(6, 7) + P(VP \rightarrow VBD NP) \cdot \beta_{VBD}(3, 3) \cdot \beta_{NP}(4, 7)$$

$$= 0.6 \cdot 1.0 \cdot 0.3 + 0.4 \cdot 1.0 \cdot 0.015 = 0.186$$
Different Parses

• Consider
  – Different splitting points:
    *E.g.,* 5th and 3rd position
  – Using different rules for VP expansion:
    *E.g.,* VP $\rightarrow$ VP PP, VP $\rightarrow$ VBD NP

• Different parses for the VP “sprayed the building with bullets” can be constructed this way.
Outside Probabilities $\alpha_j(p,q)$

**Base case:**

$\alpha_1(1,m) = 1$ for start symbol

$\alpha_j(1,m) = 0$ for $j \neq 1$

**Inductive step for calculating $\alpha_j(p,q)$:**

$\alpha_f(p,e)$

$P(N^f \rightarrow N^j N^g)$

$\beta_g(q+1,e)$

Summation over $f, g \& e$
Probability of a Sentence

- Joint probability of a sentence $w_{1m}$ and that there is a constituent spanning words $w_p$ to $w_q$ is given as:

$$P(w_{1m}, N_{pq} \mid G) = \sum_j P(w_{1m} \mid N^j_{pq}, G) = \sum_j \alpha_j(p, q)\beta_j(p, q)$$

$$P(\text{The gunman...bullets, } N_{4,5} \mid G)$$

$$= \sum_j P(\text{The gunman...bullets} \mid N^j_{4,5}, G)$$

$$= \alpha_{NP}(4, 5)\beta_{NP}(4, 5)$$

$$+ \alpha_{VP}(4, 5)\beta_{VP}(4, 5) + \ldots$$