Probabilistic Context Free Grammars (PCFGs)

We assume a fixed set of nonterminal symbols (e.g., syntactic categories) and terminal symbols (individual words). We let $X, Y,$ and $Z$ range over nonterminals and $w$ range over terminals.

A rule has the form

$$X \rightarrow \alpha_1 \alpha_2 \ldots \alpha_n$$

where $\alpha_i$ can be either a terminal or nonterminal symbol.

A grammar has a distinguished start nonterminal (the sentence category) and a set of rules.

A grammar determines a language (a set of terminal strings).

1 Chomsky Normal Form

We can always replace $\alpha_i \alpha_{i+1}$ by a fresh nonterminal $Y$ and the rule $Y \rightarrow \alpha_i \alpha_{i+1}$.

If $\alpha_i$ is a terminal symbol we can replace it by a new nonterminal $Y$ and the rule $Y \rightarrow \alpha_i$.

By repeating these transformations we get a grammar in Chomsky normal form where all productions are in one of the two forms $X \rightarrow YZ$ or $X \rightarrow w$.

2 CKY Parsing

Suppose we are given a string $w_1 w_2 \cdots w_T$

$$\text{Chart}[X, i, i] = \begin{cases} 1 & \text{if } X \rightarrow w_i \text{ is a rule} \\ 0 & \text{otherwise} \end{cases}$$

$$\text{Chart}[X, i, j] = \bigvee_{ji \leq k < j} \bigvee_{X \rightarrow YZ} \text{Chart}(Y, i, k) \wedge \text{Chart}[Z, k + 1, j]$$
3 Probabilistic Context Free Grammars

A PCFG assigns each rule $X \rightarrow \beta$ a probability $P[X \rightarrow \beta]$ satisfying the following.

$$\sum_{\beta} P[X \rightarrow \beta] = 1$$

A PCFG assigns a probability to each string.

4 Parsing Chomsky Normal Form PCFGs: The Inside Algorithm

Consider a string $w_1 \ldots w_n$. Let Chart[$X, i, j]$ be the probability that $X$ generates $w_i \ldots w_j$.

$$\text{Chart}[X, i, i] = P[X \rightarrow w_i]$$

$$\text{Chart}[X, i, j] = \sum_{k: i \leq k < j} \sum_{X \rightarrow YZ} P[X \rightarrow YZ] \text{Chart}(Y, i, k)\text{Chart}(Z, k + 1, j)$$

5 Inside Running Time

$$\text{Chart}[X, i, i] = P[X \rightarrow w_i]$$

for $2 \leq l \leq n$

for $1 \leq i \leq n - l + 1$

for $i < k < i + l$

for $X \rightarrow YZ$

$$\text{Chart}[X, i, j] + = P[X \rightarrow YZ] \text{Chart}(Y, i, k)\text{Chart}(Z, k + 1, j)$$
The inside algorithm is \( O(|G|^3) \) where \(|G|\) is the number of productions in the grammar.

6 Consistency

It is not necessarily true that a PCFG yields a distribution over strings. The sum over all strings of the probability of that string can be strictly less than one. This happens if there a nonzero probability that the expansion process fails to terminate. consider

\[
P(S \rightarrow w) = p
\]
\[
P(S \rightarrow SS) = 1 - p
\]

The number of occurrences of \( S \) in the expansion forms a random walk. If \( p < 1/2 \) there is a nonzero probability that this walk never reaches zero.

7 A Test for Consistency

Let \( M_{i,j} \) be the probability that when \( X_i \) is expanded, the expansion includes \( X_j \). A PCFG is consistent if all eigenvalues of \( M \) have magnitude less than one (the spectral radius is less than one). If some eigenvalue has magnitude greater than one, and all nonterminals are reachable from the sentence symbol, the PCFG is inconsistent.

8 Consistency for Counts

Consider a sample \( S \) of derivation trees, e.g., the Penn treebank. Let \( P_S(X \rightarrow \gamma) \) be \( \text{count}(X \rightarrow \gamma)/\text{count}(X) \) where \( \text{count}(X) \) is the number of occurrences of \( X \) in the derivation trees and \( \text{count}(X \rightarrow \gamma) \) is the number of occurrences of the production \( X \rightarrow \gamma \).

\( P_S \) is the maximum likelihood PCFG for the sample \( S \), i.e., it maximizes \( P(S|G) \) over PCFGs \( G \). \( P_S \) is always consistent.
9 Problem

Show that any PCFG can be put in Chomsky normal form in such a way that it defines the same probability for each string.